GENERALIZED PERMUTOHEDRA FROM PROBABILISTIC GRAPHICAL MODELS

FATEMEH MOHAMMADI, CAROLINE UHLER, CHARLES WANG, AND JOSEPHINE YU

Abstract. A graphical model encodes conditional independence relations via the Markov properties. For an undirected graph these conditional independence relations can be represented by a simple polytope known as the graph associahedron, which can be constructed as a Minkowski sum of standard simplices. There is an analogous polytope for conditional independence relations coming from a regular Gaussian model, and it can be defined using multiinformation or relative entropy. For directed acyclic graphical models and also for mixed graphical models containing undirected, directed and bidirected edges, we give a construction of this polytope, up to equivalence of normal fans, as a Minkowski sum of matroid polytopes. Finally, we apply this geometric insight to construct a new ordering-based search algorithm for causal inference via directed acyclic graphical models.

1. Introduction

A graphical model encodes conditional independence (CI) relations via the Markov properties. Our main goal is to understand the polyhedral geometry and combinatorics of the collection of CI relations encoded by a directed acyclic graph (DAG), a directed graph without directed cycles. It is natural, especially in view of causal inference, to associate to each conditional independence statement a collection of pairs of adjacent permutations of random variables that are compatible with that statement. Each of these pairs can be viewed as an edge of a permutohedron or a wall in the S_n fan, which is the normal fan of the permutohedron. Removing these walls gives a coarsening of the fan and a natural question is whether this fan is the normal fan of a polytope.

For undirected graphical models, the theory is well-understood. The coarsening of the S_n fan corresponding to the CI relations encoded by an undirected graph is the normal fan of a polytope called a *graph associahedron* [MPS+09]. These polytopes are Minkowski sums of standard simplices, and their facial structure has a nice description via *tubings* [CD06, PRW08].

In this paper we will show that the coarsened S_n fan arising from any DAG model is the normal fan of a polytope, which we call a *DAG associahedron*. We give two concrete constructions of DAG associahedra, one using multiinformation or relative entropy, and another using matroids. While in this paper we mainly concentrate on DAG models, we also show that these two constructions can be extended to more general graphical models that have been studied in the literature containing a mix of undirected, directed and bidirected edges. In contrast to graph associahedra, we show that DAG associahedra are in general not simple polytopes and cannot be realized as a Minkowski sum of standard simplices. Our main motivation for studying DAG associahedra is causal inference: Given a set of CI relations that are inferred from data, the goal is to estimate the underlying DAG model, also known as a Bayesian network. A DAG is defined by an ordering of the nodes and an undirected graph. We show how our geometric insight on DAG associahedra can be applied to construct a new ordering-based search algorithm for causal inference.

Keywords: Graphical model, graphoid, permutohedron, causal inference, submodular function, matroid, entropy.
MSC(2010): 62H05 (primary); 52B12, 52B40 (secondary).

Other polyhedral approaches for learning Bayesian networks have been described in the literature [CHS16, HLS12, JSGM10, SVH10, SHHL12]. These approaches are based on using integer programming or linear programming relaxations to maximize a score function over a polytope; most notably, the Family Variable Polytope (FVP) and the Characteristic Imset Polytope (CIP), whose vertices correspond to all possible DAGs on n nodes, up to Markov equivalence, respectively. While the FVP and the CIP are high-dimensional ($n(2^{n-1}-1)$ and (2^n-n-1), respectively) and very complex polytopes (facet description only known for $n \leq 4$), we here present a new polyhedral approach for learning Bayesian networks that is based on DAG associahedra, $n-1$-dimensional polytopes for which we give a concrete construction.

2. Notation and background

In this section, we discuss the relationship between CI relations, the S_n fan, and generalized permutohedra. Please refer to Appendix A for basic definitions of polytopes and fans and Appendix C for a "dictionary" of concepts.

Let $[n] = \{1, \ldots, n\}$, and let $\mathbb{P}$ be a joint distribution on the random variables X_i for $i \in [n]$. For notational simplicity we often write I for $\{X_i : i \in I\}$ where $I \subseteq [n]$. For pairwise disjoint subsets $I, J, K \subset [n]$ we say that I is *conditionally independent* of J given K under $\mathbb{P}$ if the conditional probability $\mathbb{P}(\mathcal{A} \mid J, K)$ does not depend on J for any measurable set $\mathcal{A}$ in the sample space of X_I. This statement is denoted by $I \perp\!\!\!\perp_{\mathbb{P}} J \mid K$ or simply $I \perp\!\!\!\perp J \mid K$. If $K = \varnothing$, we write $I \perp\!\!\!\perp J$. The set of CI relations arising from a distribution satisfies the following basic implications, known as the *semigraphoid* properties [Pea88]:

(SG1') if $I \perp\!\!\!\perp J \mid L$ then $J \perp\!\!\!\perp I \mid L$,
(SG2') if $I \perp\!\!\!\perp J \mid L$ and $U \subseteq I$, then $U \perp\!\!\!\perp J \mid L$,
(SG3') if $I \perp\!\!\!\perp J \mid L$ and $U \subseteq I$, then $I \perp\!\!\!\perp J \mid (U \cup L)$,
(SG4') if $I \perp\!\!\!\perp J \mid L$ and $I \perp\!\!\!\perp K \mid J \cup L$, then $I \perp\!\!\!\perp (J \cup K) \mid L$.

In this paper, CI relations can be considered as a formal construct and do not necessarily need any probabilistic interpretation. In addition, we will only work with relations in which I and J are both singletons, denoted by lower-case letters i, j. To simplify notation, we use concatenation to denote union among subsets and elements of $[n]$, e.g. Lij means $L \cup \{i, j\}$. Then a *semigraphoid* is a set of CI relations that satisfy the CI implications

(SG1) if $i \perp\!\!\!\perp j \mid L$ then $j \perp\!\!\!\perp i \mid L$,
(SG2) if $i \perp\!\!\!\perp j \mid L$ and $i \perp\!\!\!\perp k \mid jL$, then $i \perp\!\!\!\perp k \mid L$ and $i \perp\!\!\!\perp j \mid kL$,

for distinct $i, j, k \in [n]$ and $L \subseteq [n] \setminus \{i, j, k\}$.

For distributions with strictly positive densities such as regular Gaussian distributions, the *intersection axiom* holds in addition to the semigraphoid axioms, namely

(INT) if $i \perp\!\!\!\perp j \mid kL$ and $i \perp\!\!\!\perp k \mid jL$, then $i \perp\!\!\!\perp j \mid L$ and $i \perp\!\!\!\perp k \mid L$.

The implications (SG1), (SG2) and (INT) together are known as the *graphoid* properties. Note that these implications are not a complete list of CI implications that hold for distributions. In fact, Studený [Stu92] proved that there exists no finite such characterization.

In [LM07], Lněnička and Matúš defined *gaussoids* as the graphoids satisfying the following additional axioms:

(G1) if $i \perp\!\!\!\perp j \mid L$ and $i \perp\!\!\!\perp k \mid L$, then $i \perp\!\!\!\perp j \mid kL$ and $i \perp\!\!\!\perp k \mid jL$,
(G2) if $i \perp\!\!\!\perp j \mid L$ and $i \perp\!\!\!\perp j \mid kL$, then $i \perp\!\!\!\perp k \mid L$ or $j \perp\!\!\!\perp k \mid L$.

The property (G1) is the converse of the intersection axiom, and (G2) is known as *weak transitivity.* The CI relations of any regular Gaussian distribution form a gaussoid, but not all gaussoids arise this way. The set of CI relations coming from any undirected graphical model or a DAG model can be faithfully represented by a regular Gaussian distribution, hence forming a gaussoid.

We will associate a geometric object to a collection of CI relations as follows: Consider the hyperplanes in $\mathbb{R}^n$ defined by equations of the form $x_i = x_j$ for all $1 \le i < j \le n$. The complement of these hyperplanes consists of points in $\mathbb{R}^n$ with distinct coordinates, and they are partitioned into $n!$ connected components corresponding to the permutations of $[n]$ as follows: We identify a permutation (bijection) $\pi : [n] \to [n]$ with the linear order $\pi(1) \succ \pi(2) \succ \cdots \succ \pi(n)$. To every vector $u \in \mathbb{R}^n$ with distinct coordinates, we associate a linear order $\succ$ on $[n]$ by defining $i \succ j$ if and only if $u_i > u_j$. For example, the vector $u = (25, 4, 16, 9)$ gives the linear order $1 \succ 3 \succ 4 \succ 2$, which we denote using its *descent vector* of the form $(1|3|4|2)$. Two points in the complement of the hyperplanes $x_i = x_j$ in $\mathbb{R}^n$ are in the same connected component if and only if they have the same descent vector.

The closures of the $n!$ cones and all their faces form a fan, which we will call the *S_n fan.* It is also known as the *permutohedral fan* or the *A_{n-1} fan* or the *braid arrangement fan.* Each cone in the fan contains the line in direction $(1, 1, \ldots, 1)$ and is generated by a collection of 0/1 vectors, every pair of which is nested (when each 0/1 vector is identified with its set of non-zero coordinates).

To each CI relation $i \perp\!\!\!\perp j \mid K$, where $i, j \in [n]$ distinct and $K \subseteq [n] \setminus \{i, j\}$, we associate pairs of adjacent permutations of the form

$$(a_1|\cdots|a_k|i|j|b_1|\cdots|b_{n-k-2}) \text{ and } (a_1|\cdots|a_k|j|i|b_1|\cdots|b_{n-k-2}), \tag{1}$$

where $\{a_1, \ldots, a_k\} = K$ and $\{b_1, \ldots, b_{n-k-2}\} = [n]\backslash(K \cup \{i, j\})$. We will denote such a pair by $(a_1|\cdots|a_k|i\,j|b_1|\cdots|b_{n-k-2})$. For each relation $i \perp\!\!\!\perp j \mid K$ there are $|K|!\,(n - |K| - 2)!$ such pairs.

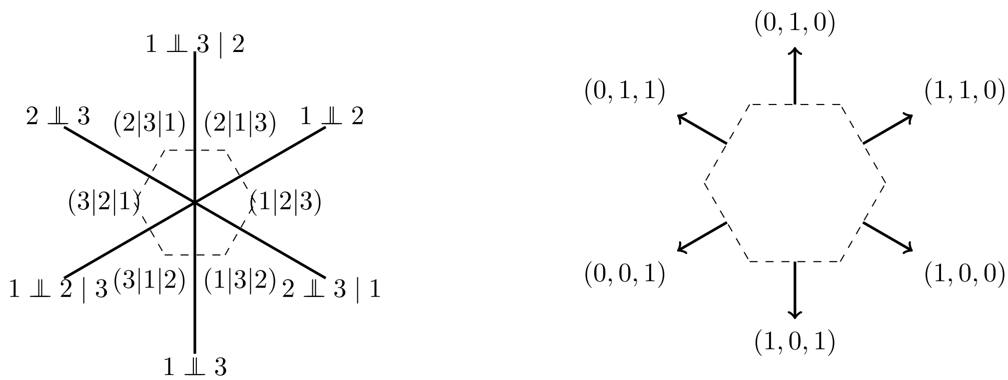
A fan F in $\mathbb{R}^n$ is said to be a *coarsening* of the S_n fan if every cone in the S_n fan is contained in a cone of F, or equivalently, if every cone of F is a union of some cones of the S_n fan. In particular, maximal cones of F are unions of maximal cones of the S_n fan, and F can be constructed from the S_n fan by removing certain walls (codimension one cones). This gives an equivalence relation on S_n — two permutations are equivalent if and only if their corresponding cones in the S_n fan are contained in the same cone in F. Such an equivalence relation coming from a fan is called a *convex rank test* in [MPS+09]. We will see in §8 that for DAG models this equivalence relation coincides with that coming from the Sparsest Permutation Algorithm of Raskutti and Uhler [RU14].

We identify a coarsening of the S_n fan with the collection of walls that are removed. Each wall corresponds to an adjacent pair of permutations as in (1), which gives a CI relation $i \perp\!\!\!\perp j \mid \{a_1, \ldots, a_k\}$. It was shown in [MPS+09, Theorem 6] that a set of walls form the missing walls in a fan that coarsens the S_n fan if and only if the corresponding set of CI relations forms a semigraphoid. In particular, if the wall associated to the pair (1) is not a wall in a coarsened S_n fan F, then any pair obtained by permuting the a's among themselves and b's among themselves is also not a wall in F.

A complete fan F in $\mathbb{R}^n$ is called *polytopal* if it is the normal fan of a polytope. The S_n fan itself is polytopal since it is the normal fan of a *permutohedron* P_n defined as follows. Let $a_1 < a_2 < \cdots < a_n$ be real numbers. Let

$$P_n = \operatorname{conv}\{(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}) : \sigma \in S_n\} \subset \mathbb{R}^n.$$

Different choices of a_i's give different polytopes but with the same normal fan. We associate to each vertex of P_n a permutation given by its descent vector as explained above, e.g. a point with coordinates $(2, 3, 4, 1) \in \mathbb{R}^4$ is associated with its descent vector $(3|2|1|4)$, which is a permutation and *not* a point in $\mathbb{R}^4$. Two vertices of P_n are connected by an edge if and only if their descent vectors differ by an adjacent transposition as in (1). Thus each CI relation corresponds to a certain set of edges of P_n.



(a) The S_3 fan modulo the line $(1, 1, 1)$. Maximal cones are labeled with permutations and the walls are labeled with CI relations.

(b) Permutohedron P_3 with outer normals of its facets.

FIGURE 1. The permutohedron P_3 and its normal S_3-fan. Only the descent vectors, not coordinate vectors, of the vertices of P_3 are shown in (a).

A *generalized permutohedron* (see [PRW08]) is a polytope whose normal fan is a coarsening of the S_n fan. See Figures 1 and 2 for some examples. These polytopes have other equivalent definitions, and are also called *M-convex polytopes* or *base polyhedra* [Mur03, (4.43)]. Their projections along a coordinate direction give *generalized polymatroids* [Fuj05, Theorem 3.58]. We use the term generalized permutohedron to highlight the connection to permutations.

Example 2.1 (Undirected graphical models and graph associahedra)**.** Let G be an undirected graph with node set $[n]$. We associate a random variable X_i to each node i of the graph. The joint distribution $\mathbb{P}$ of the random vector $X = (X_1, \ldots, X_n)$ satisfies the *undirected (global) Markov property* with respect to G if $I \perp\!\!\!\perp J \mid K$ for all disjoint subsets $I, J, K \subset [n]$ such that K *separates* I and J in G, i.e. every path between nodes $i \in I$ and $j \in J$ passes through a node $k \in K$. If a distribution $\mathbb{P}$ satisfies exactly the CI relations corresponding to separations in the graph G, then $\mathbb{P}$ is called *faithful* or *perfectly Markovian* with respect to G.

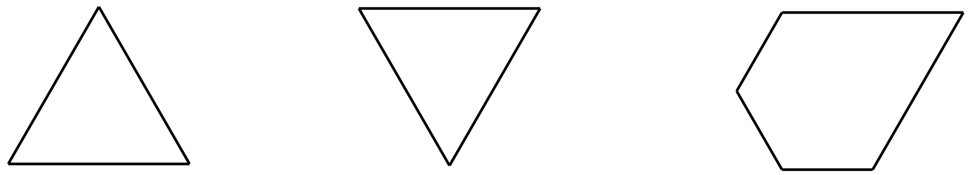
For any undirected graph there exist faithful regular Gaussian distributions; see [Lau96, Chapter 3] for more details. Hence for any undirected graph G the corresponding CI relations defined by the Markov property satisfy the gaussoid axioms. The coarsened S_n fan associated to the gaussoid of an undirected graph is the normal fan of a polytope, which can be realized as the Minkowski sum of standard simplices $\Delta_I = \mathrm{conv}\{e_i : i \in I\}$ where I runs over all sets of nodes that induce connected subgraphs of G [MPS+09]. These polytopes are called *graph associahedra* and were studied in [Dev09, CD06, PRW08]. □

We will now summarize a characterization of coarsened S_n fans that are polytopal, based on [MPS+09] and [HMS+08]. Let $2^{[n]}$ denote the power set of $[n]$, the set of all subsets of $[n]$. A function $\omega : 2^{[n]} \to \mathbb{R}$ is called *submodular* if

$$\omega(Ki) + \omega(Kj) \geq \omega(Kij) + \omega(K) \qquad (2)$$

for all $K \subset [n]$ and $i, j \in [n] \backslash K$. A submodular function also satisfies $\omega(A) + \omega(B) \geq \omega(A \cup B) + \omega(A \cap B)$ for all $A, B \subset [n]$. Note that a submodular function on $2^{[n]}$ is an *L-convex* function on the unit cube $\{0, 1\}^n$ [Mur03].

Definition 2.2. A semigraphoid on $[n]$ is called *submodular* if there is a submodular function ω on $2^{[n]}$ with $\omega(\varnothing) = 0$ such that $\omega(Ki) + \omega(Kj) = \omega(Kij) + \omega(K)$ if and only if the relation $i \perp\!\!\!\perp j \mid K$ is in the semigraphoid.



(a) The standard simplex $\mathrm{conv}\{e_1, e_2, e_3\}$.

(b) The matroid polytope of $U_{2,3}$, $\mathrm{conv}\{e_1 + e_2, e_1 + e_3, e_2 + e_3\}$.

(c) The DAG associahedron of $1 \to 3 \leftarrow 2$ with CI relation $1 \perp\!\!\!\perp 2$. This polytope is not a Minkowski sum of scaled standard simplices.

FIGURE 2. Some generalized permutohedra. Compare with the fan in Figure 1.

Submodular semigraphoids correspond to *structural independence models* [Stu05, §5.4.2], which can be viewed as semigraphoids obtained from supermodular functions, whose negatives are submodular functions.

The following result shows that every submodular function determines a semigraphod, and the semigraphoids arise this way are precisely those corresponding to polytopal coarsenings of the S_n fan.

Lemma 2.3. *A polytope $P \subset \mathbb{R}^n$ is a generalized permutohedron if and only if there exists a submodular function $\omega : 2^{[n]} \to \mathbb{R}$ with $\omega(\varnothing) = 0$ such that*

$$P = \{x \in \mathbb{R}^n : \sum_{i \in I} x_i \leq \omega(I) \text{ for each non-empty } I \subset [n], \text{ and } \sum_{i \in [n]} x_i = \omega([n])\}. \tag{3}$$

A wall in the S_n fan corresponding to $i \perp\!\!\!\perp j \mid K$ is missing in the normal fan of P defined by ω as above if and only if $\omega(Ki) + \omega(Kj) = \omega(Kij) + \omega(K)$. In particular, a coarsened S_n fan is polytopal if and only if the corresponding semigraphoid is submodular.

The lemma follows from the conjugacy between L- and M- convex functions and also from [Mur03, Theorem 4.15]. A part of it appeared in [MPS+09, Proposition 12 and Theorem 14]. We provide a proof in Appendix B, as it is difficult to find a complete proof in the literature.

Remark 2.4. *If ω is a submodular function on $2^{[n]}$ with $\omega(\varnothing) = 0$, then $\omega' : 2^{[n]} \to \mathbb{R}$ defined as $\omega'(S) = \omega([n] \backslash S) - \omega([n])$ is also submodular with $\omega'(\varnothing) = 0$. The polytopes P and P', defined by ω and ω' as in (3), are related by $-P = P'$.* □

Example 2.5. Consider the submodular function ω on $2^{[n]}$ whose value is 1 on all non-empty sets and 0 on the empty set. This function is known as the rank function of the uniform rank one matroid on $[n]$. The generalized permutohedron defined by this submodular function is a standard simplex of dimension $n-1$ whose outer normal vectors are e_I for subsets I of size $n-1$. Any set of $n-2$ facet normals spans a wall in the normal fan, with pairs of the form (1), where $K = \varnothing$, corresponding to relations of the form $i \not\perp\!\!\!\perp j \mid \varnothing$. See Figures 1 and 2. □

This characterization leads to the following questions for any given semigraphoid:

Question A. *Is a given semigraphoid submodular? And if so, can we construct a submodular function with the desired equalities as in Definition 2.2?*

In the following sections we will give a positive answer to these questions for semigraphoids coming from DAG models.

Rank functions of matroids are submodular functions, so every matroid M on the ground set $[n]$ gives a semigraphoid on $[n]$ as follows:

$$i \not\perp\!\!\!\perp j \mid K \iff \mathrm{rank}(Ki) + \mathrm{rank}(Kj) > \mathrm{rank}(Kij) + \mathrm{rank}(K)$$

Note that since a matroid rank function takes integer values and $\mathrm{rank}(Aa) \leq \mathrm{rank}(A) + 1$ for any $A \subset [n]$ and $a \in [n]$, we obtain

$$i \not\perp\!\!\!\perp j \mid K \iff \mathrm{rank}(K) + 1 = \mathrm{rank}(Ki) = \mathrm{rank}(Kj) = \mathrm{rank}(Kij). \tag{4}$$

In this case, the coarsening of the S_n fan is the outer normal fan of the *matroid polytope*, which is defined as the convex hull of the indicator functions of the bases of the matroid. For example, the standard simplex $\Delta_I = \mathrm{conv}\{e_i : i \in I\}$ is the matroid polytope of the rank one matroid in which each element of I forms a base. The intersection of two semigraphoids (as sets of conditional independence relations) is again a semigraphoid. This intersection operation corresponds to common refinement, Minkowski sum, and sum, respectively, for fans, polytopes, and submodular functions.

Question B. *Which submodular semigraphoids can be obtained from the sums of rank functions of matroids (as in Definition 2.2)? Which fans arising from semigraphoids are normal fans of Minkowski sums of matroid polytopes?*

For example, the Minkowski sum of all standard simplices Δ_I, for all non-empty subsets $I \subset [n]$, is affinely equivalent to the permutohedron P_n, i.e. they have the same normal fan, which is the entire S_n fan. This decomposition is not unique, however, e.g. P_3 is a hexagon and can be decomposed as the Minkowski sum of either two triangles or three line segments, all of which are matroid polytopes.

Example 2.6. Let $\mathcal{G}$ be the following DAG.

$$\begin{tikzcd} 1 \arrow[dr] & & 2 \arrow[dl] \\ & 3 & \end{tikzcd}$$

We will see in the next section that the Markov property on $\mathcal{G}$ defines a single CI relation, namely $1 \perp\!\!\!\perp 2$. Removing the corresponding wall in the S_3 fan gives a fan with 5 maximal cones. Figure 2(c) depicts a polytope with this normal fan. It is straightforward to check that this fan is not the normal fan of a Minkowski sum of standard simplices, but it is the normal fan of the Minkowski sum of the simplex in Figure 2(b) together with two line segments, which are all matroid polytopes. □

3. Bayesian networks

Similarly to undirected graphs we can define probabilistic models on DAGs. Such graphical models are also known as *Bayesian networks*.

Let $\mathcal{G}$ be a DAG with nodes $[n]$. If there is a directed edge from i to j in $\mathcal{G}$, which we denote by $i \to j$ in $\mathcal{G}$ or $(i, j) \in \mathcal{G}$, the node i is called a *parent* of the node j. The set of all parent nodes of j is denoted by *pa(j)*.

We now review the concept of separation for DAGs. A *path* in $\mathcal{G}$ is an alternating sequence of nodes and edges, starting and ending at nodes, in which each edge is adjacent in the sequence to its two endpoints[1]. The path may contain repeated edges and nodes. We do *not* assume that the direction of the edges is compatible with the ordering of the nodes in the path.

[1] This is often called a "walk", but we prefer to use "path" in order to be consistent with the notion of a "Bayes ball path".

Definition 3.1. Let $\mathcal{G}$ be a DAG on $[n]$ and let $i, j \in [n]$ and $K \subset [n] \setminus \{i, j\}$. A *Bayes ball path from i to j given K* in $\mathcal{G}$ is a path from i to j in $\mathcal{G}$ such that

(1) if $a \to b \to c$ or $a \leftarrow b \to c$ or $a \leftarrow b \leftarrow c$ is on the path, then $b \notin K$;
(2) if $a \to b \leftarrow c$ is on the path, then $b \in K$ (where a and c need not be distinct). In this case the node b is called a *collider* along the path.

See Figures 3 and 5 for examples. Informally we think of a directed edge $i \to j$ as pointing *down* from i to j. A "Bayes ball" rolls along edges of the DAG. It cannot roll through nodes that are in K, but it can "bounce off" them by going down, touching K, then going back up either along the same or a different edge.

For subsets of nodes $I, J, K \subset [n]$, we say that I and J are *directionally separated* or *d-separated* by K in $\mathcal{G}$ if there is no Bayes-ball path from any element of I to any element of J given K [VP90]. This led to the construction of the *Bayes-Ball algorithm* [Sha98], an algorithm for determining d-separation statements. Similarly as for undirected graphs, we can also associate a random vector with joint distribution $\mathbb{P}$ to the nodes of a DAG $\mathcal{G}$. Then $\mathbb{P}$ satisfies the *directed (global) Markov property* with respect to $\mathcal{G}$ if $I \perp\!\!\!\perp J \mid K$ for all disjoint subsets $I, J, K \subset V$ such that K d-separates I and J in $\mathcal{G}$. A faithful distribution to $\mathcal{G}$, i.e. a distribution that satisfies exactly the CI relations corresponding to d-separation in $\mathcal{G}$, can be realized by regular Gaussian distributions (see §4). Hence, for any DAG $\mathcal{G}$ the CI relations of the form $i \perp\!\!\!\perp j \mid K$, where i and j are d-separated given K in $\mathcal{G}$, form a gaussoid, which we call a *DAG gaussoid*.

It is important to note that while the set of separation statements uniquely determines an undirected graph, this is not the case for d-separation statements for DAGs. Two DAGs are called *Markov equivalent* if they imply the same d-separation statements. The Markov equivalence class is determined by the skeleton of a DAG and its *V-structures* — triples of nodes (i, j, k) such that $i \to k \leftarrow j$ and i, j are not adjacent [AMP97]. An *essential graph* [AMP97] (also called a *completed partially directed acyclic graph* or *CPDAG* in [Chi02a] and a *maximally oriented graph* in [Mee95]) is a graph with undirected and directed edges that uniquely represents a Markov equivalence class of DAGs. It has the same skeleton as the DAGs in the Markov equivalence class and contains a directed edge $i \to j$ if and only if each DAG in the Markov equivalence class contains the directed edge $i \to j$.

The following is our main result and answers Questions A and B for DAG gaussoids.

Theorem 3.2 (Main Theorem)**.** *Every DAG gaussoid is submodular. Equivalently, the associated coarsening of the S_n fan is the normal fan of a polytope. Moreover, there is a realization of this polytope as a Minkowski sum of matroid polytopes.*

The equivalence of the first two statements was proven in Lemma 2.3 above. We call any such polytope resulting from a DAG gaussoid a *DAG associahedron*. DAG associahedra are uniquely defined up to equivalence of normal fans, and they only depend on the DAG up to Markov equivalence.

Remark 3.3. *Let $\mathcal{G}$ be a DAG. The normal fan of the DAG associahedron corresponding to $\mathcal{G}$ can be obtained by coarsening the normal fan of the graph associahedron corresponding to the* moral graph *of $\mathcal{G}$ — the undirected graph with edges (i, j) if $i \to j$ in $\mathcal{G}$, $j \to i$ in $\mathcal{G}$, or $i \to k \leftarrow j$ for some k in $\mathcal{G}$; see Figure 3.* $\square$

In the next two sections, we will give two independent proofs for the submodularity of DAG gaussoids. In the first proof, in §4, we use multiinformation, or relative entropy, to give a formula for the submodular function and hence a realization of DAG associahedra. However, in general the constant terms of the inequalities in this construction are not rational. We will discuss some

Figure 3. The DAG $\mathcal{G}$ (left) and its moral graph G (right) discussed in Example 3.4.

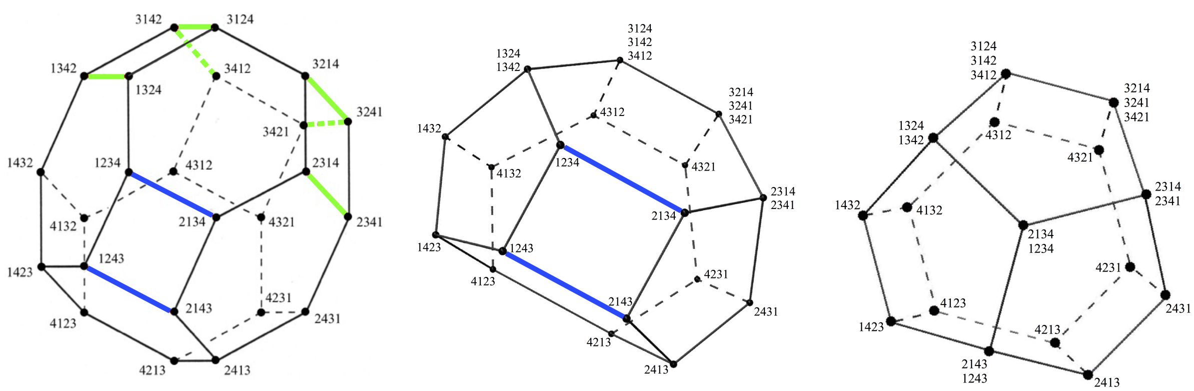
heuristic methods for finding exact combinatorial information from approximate inequalities. In the second proof, in §5, we give a realization of DAG associahedra as Minkowski sums of matroid polytopes, which are integral polytopes. The submodularity of a semigraphoid can be tested using linear programming [HMS+08]. So our theorem states that the linear programs coming from DAG gaussoids are always feasible, and our proofs give an explicit construction of a feasible solution.

We illustrate the concepts introduced so far by an example of a DAG model on 4 nodes and describe the corresponding DAG associahedron.

Example 3.4. Consider the DAG $\mathcal{G}$ shown in Figure 3. An example of a Bayes ball path in $\mathcal{G}$ is the path from node 1 to 2 given $K = \{4\}$, since on the path $1 \to 3 \to 4 \leftarrow 3 \leftarrow 2$ the node $3 \notin K$ but $4 \in K$. The DAG gaussoid corresponding to $\mathcal{G}$ consists of the CI relations

$$1 \perp\!\!\!\perp 2, \quad 1 \perp\!\!\!\perp 4 \mid 3, \quad 2 \perp\!\!\!\perp 4 \mid 3, \quad 1 \perp\!\!\!\perp 4 \mid \{2,3\}, \quad 2 \perp\!\!\!\perp 4 \mid \{1,3\}.$$

The corresponding edges of the permutohedron are shown in green and blue in Figure 4(a). Since these CI relations form a semigraphoid, we obtain a coarsening of the S_n fan by removing the edges (12|3|4), (12|4|3), (3|14|2), (3|24|1), (2|3|14), (3|2|14), (1|3|24) and (3|1|24). The resulting coarsening of the S_n fan obtained by contracting the colored edges in the permutohedron is polytopal. The convex polytope corresponding to this DAG associahedron is shown in Figure 4(c).



(a) Permutohedron P_4. The green edges correspond to CI relations in the moral graph G in Example 3.4. The blue edges correspond to the additional CI relations in $\mathcal{G}$.

(b) The graph associahedron of the moral graph G obtained by contracting the green edges.

(c) DAG associahedron for $\mathcal{G}$ obtained by contracting both, green and blue edges.

Figure 4. The vertices are labeled by descent vectors of permutations, with "|"s removed. The figures show how the combinatorics of the polytope changes as edges are contracted, but the polytopes are not drawn to be geometrically correct.

The moral graph G of $\mathcal{G}$ is shown in Figure 3 (right). The gaussoid corresponding to G consists of the CI relations

$$1 \perp\!\!\!\perp 4 \mid 3, \quad 2 \perp\!\!\!\perp 4 \mid 3, \quad 1 \perp\!\!\!\perp 4 \mid \{2,3\}, \quad 2 \perp\!\!\!\perp 4 \mid \{1,3\}.$$

In general, any DAG gaussoid contains the gaussoid of its moral graph. The edges corresponding to the CI relations for the moral graph are shown in green in Figure 4(a). By contracting the green edges in the permutohedron we obtain the graph associahedron corresponding to G shown in Figure 4(b). By further contracting also the blue edges, we obtain the DAG associahedron corresponding to $\mathcal{G}$.

As we will see in Proposition 3.6, the DAG associahedron in this example cannot be realized as a Minkowski sum of simplices. However, we will show in §5 that it can be realized as the following Minkowski sum of matroid polytopes:

$$\Delta_{13} + \Delta_{23} + \Delta_{34} + \Delta_{134} + \Delta_{234} + \mathrm{conv}\{e_{12}, e_{13}, e_{23}\} + \mathrm{conv}\{e_{12}, e_{13}, e_{23}, e_{14}, e_{24}\}.$$

As we will see in §5, the first three polytopes in the sum correspond to the three edges in the DAG, the next two correspond to the paths 1 3 4 and 2 3 4, which have no colliders, and the last two correspond to the paths 1 3 2 (given 3) and 1 3 4 3 2 (given 4) respectively. See Example 5.4. □

We end this section with two observations about DAG associahedra. In the following example we show that unlike graph associahedra, DAG associahedra need not be *simple*.

Example 3.5 (A non-simple DAG associahedron)**.** Let $\mathcal{G}$ be the following DAG.

$$\begin{tikzcd} 1 \arrow[r] \arrow[d] & 2 \arrow[d] \arrow[dl] \\ 4 & 3 \arrow[l] \end{tikzcd}$$

The corresponding DAG gaussoid consists only of the CI relation $1 \perp\!\!\!\perp 3 \mid 2$, which corresponds to a single edge $2|13|4$ on the permutohedron P_4. Contracting this edge gives a vertex adjacent to 4 edges on a 3-dimensional polyhedron, so the resulting polytope is not simple. In this case, the combinatorial operation of contracting the edge can be realized geometrically by pushing the two neighboring square facets towards each other until they meet at a vertex. In other words, the edge shared by two hexagonal faces contracts to a single vertex. □

Furthermore, as already mentioned in Example 3.4, unlike graph associahedra, DAG associahedra need not be Minkowski sums of standard simplices (MSS) in the sense of [MPS$^+$09]. In fact, the following result shows that a DAG associahedron can be realized as a MSS if and only if the DAG gaussoid equals the gaussoid of its moral graph, or in other words, if and only if the DAG model coincides with an undirected graphical model.

Proposition 3.6. *The DAG associahedron associated to a DAG $\mathcal{G}$ is MSS if and only if $\mathcal{G}$ does not contain any V-structures, i.e. the DAG model coincides with an undirected graphical model.*

We saw in Example 2.6 that a V-structure cannot be MSS. We can generalize this example to the following corollary of [MPS$^+$09, Proposition 20]:

Lemma 3.7. *If a semigraphoid arises from a Minkowski sum of standard simplices, then for any $i, j \in [n]$ distinct and $K \subseteq K' \subseteq [n] \setminus \{i,j\}$, we have*

$$i \perp\!\!\!\perp j \mid K \implies i \perp\!\!\!\perp j \mid K'.$$

Proof. For $I \subset [n]$, the standard simplex Δ_I is the matroid polytope of the rank one matroid on $[n]$ whose loops are $[n] \backslash I$. By (4) the semigraphoid corresponding to Δ_I contains $i \not\perp\!\!\!\perp j \mid K$ where

$i, j \in I$ and $K \cap I = \varnothing$. Taking a Minkowski sum of such simplices corresponds to taking the union of the associated conditional dependence statements. It follows that $i \not\perp\!\!\!\perp j \mid K' \implies i \not\perp\!\!\!\perp j \mid K$ for all $K \subseteq K' \subseteq [n] \setminus \{i, j\}$. $\square$

Using Lemma 3.7 we can now easily prove Proposition 3.6.

Proof of Proposition 3.6. If $\mathcal{G}$ does not contain any V-structures, then the corresponding DAG gaussoid is equivalent to the gaussoid obtained from an undirected graph, namely the skeleton of $\mathcal{G}$, so it is MSS.

On the other hand, suppose that $\mathcal{G}$ contains a V-structure $i \to \ell \leftarrow j$. Let $K \subset [n]$ be the set of non-descendants of i and j in $\mathcal{G}$, i.e. the set of $k \in [n]$ such that there is no directed path from i to k or from j to k in $\mathcal{G}$. Then the CI relation $i \perp\!\!\!\perp j \mid K$ is contained in the gaussoid corresponding to $\mathcal{G}$. However, the CI relation $i \perp\!\!\!\perp j \mid K\ell$ is not in the gaussoid of $\mathcal{G}$, since there is a Bayes ball path from i to j given $K\ell$ in $\mathcal{G}$. Hence by Lemma 3.7 above, the DAG associahedron corresponding to $\mathcal{G}$ is not MSS. $\square$

4. A construction of DAG associahedra from multiinformation

The *multiinformation* of a probability measure $\mathbb{P}$ on $[n]$ is a function $m_{\mathbb{P}} : 2^{[n]} \to [0, \infty]$ defined by

$$m_{\mathbb{P}}(S) = H(\mathbb{P} | \Pi_{i \in S} \mathbb{P}^{\{i\}}),$$

where H denotes the relative entropy with respect to a product of one-dimensional marginals $\mathbb{P}^{\{i\}}$. For the case of most interest to us, when $\mathbb{P}$ is a regular Gaussian, there is a simpler formula as follows. Let $\mathbb{P}$ be a regular Gaussian measure on $[n]$ with covariance matrix Σ. Let Γ be the correlation matrix of $\mathbb{P}$ — a symmetric positive definite matrix obtained from Σ by simultaneously rescaling the rows and columns so that all the diagonal entries are equal to one. In other words, $\Gamma = D^{-1/2} \Sigma D^{-1/2}$ where $D = \operatorname{diag}(\Sigma)$. Then we have

$$i \perp\!\!\!\perp j \mid K \iff \operatorname{rank}(\Gamma_{Ki, Kj}) \leq |K| \tag{5}$$

where $\Gamma_{A,B}$ denotes the submatrix of Γ with rows and columns indexed by A and B, respectively [Sul09]. By [Stu05, Corollary 2.6] the multiinformation $m_{\mathbb{P}}(A)$ for $A \subset [n]$ is

$$m_{\mathbb{P}}(A) = -\frac{1}{2} \log \det(\Gamma_{A,A}).$$

Since Γ is positive definite, all its principal minors $\det(\Gamma_{A,A})$ are non-zero. We define $\det(\Gamma_{\varnothing,\varnothing})$ to be 1. By [Stu05, Corollary 2.2] we have

$$m_{\mathbb{P}}(A) = 0 \text{ for all } A \subseteq [n], |A| \leq 1, \text{ and}$$

$$m_{\mathbb{P}}(ABC) + m_{\mathbb{P}}(C) \geq m_{\mathbb{P}}(AC) + m_{\mathbb{P}}(BC) \text{ for all } A, B, C \subset [n]$$

with equality if and only if $A \perp\!\!\!\perp B \mid C$ under $\mathbb{P}$.

We summarize this discussion in the following lemma.

Lemma 4.1. *If $\mathbb{P}$ is a regular Gaussian distribution with correlation matrix Γ, then its semigraphoid is submodular, with submodular function given by*

$$A \mapsto \log \det(\Gamma_{A,A}).$$

The submodularity of DAG gaussoids (i.e. the first part of Theorem 3.2) follows from the lemma above and the fact that any DAG gaussoid has a faithful regular Gaussian realization. See for example [DSS09, §3.3], where the following construction is described.

Let $\mathcal{G}$ be a DAG on the nodes $[n]$. Assume that the nodes are labeled so that if $i \to j$ is an edge in $\mathcal{G}$, then $i < j$. Let Λ be an upper-triangular matrix whose entries have the form

$$\Lambda_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ -\ell_{ij} & \text{if } i \to j \text{ is an edge in } \mathcal{G}, \\ 0 & \text{otherwise,} \end{cases}$$

where ℓ_{ij} are real numbers. Let $K = \Lambda\Lambda^T$ and $\Sigma = K^{-1}$. Then K is symmetric positive definite by construction, and so is Σ. For almost all choices of real numbers ℓ_{ij} (apart from an algebraic hypersurface), a Gaussian distribution $\mathbb{P}$ with covariance matrix Σ is faithful to the DAG gaussoid of $\mathcal{G}$ [URBY13].

In fact, as explained in the following lemma, the inequalities for the desired generalized permutohedron can also be computed directly from minors of $K = \Lambda\Lambda^T$ instead of from the correlation matrix $\Gamma = D^{-1/2}\Lambda^{-T}\Lambda^{-1}D^{-1/2}$, where $D = \text{diag}(\Lambda^{-T}\Lambda^{-1})$. This result simplifies computations considerably since we do not need to perform any matrix inversion on $\Lambda\Lambda^T$.

Lemma 4.2. *Let K be a positive definite matrix and let ω be the submodular function on $2^{[n]}$ given by $\omega(A) = \log\det(K_{A,A})$. Let P be the polytope defined as in (3). Then $-P$ is the generalized permutohedron corresponding to the semigraphoid of a regular Gaussian distribution $\mathbb{P}$ with covariance matrix $\Sigma = K^{-1}$.*

Proof. The polytope defined by the submodular function $A \mapsto \log\det(\Gamma_{A,A})$ is obtained from the polytope defined by the submodular function $A \mapsto \log\det(\Sigma_{A,A})$ by translation in each coordinate direction i by $-\log\Sigma_{i,i}$. Thus these two polytopes have the same normal fans and encode the same semigraphoids.

For $A \subset [n]$ and $B = [n]\backslash A$, we have $(\Sigma_{A,A})^{-1} = K_{A,A} - K_{A,B}(K_{B,B})^{-1}K_{B,A}$, the Schur complement. Using the equality $\det(K) = \det(K_{B,B}) \cdot \det(K_{A,A} - K_{A,B}(K_{B,B})^{-1}K_{B,A})$, we obtain

$$\begin{aligned} \log\det(\Sigma_{A,A}) &= -\log\det(\Sigma_{A,A})^{-1} \\ &= -\log\det(K_{A,A} - K_{A,B}(K_{B,B})^{-1}K_{B,A}) \\ &= \log\det(K_{B,B}) - \log\det(K). \end{aligned}$$

Combining this with Remark 2.4, it follows that the polytopes given by $A \mapsto \log\det(\Sigma_{A,A})$ and by $A \mapsto \log\det(K_{A,A})$ are negatives of each other. □

In other words, by using K instead of Σ we obtain the *dual semigraphoid* defined in [MPS+09]. In particular, if a semigraphoid has a faithful regular Gaussian distribution, then so does its dual.

Example 4.3 (Multiinformation of the 4-node DAG in Example 3.4)**.** We start by constructing Λ from $\mathcal{G}$ using edge weights 1 (i.e. $\ell_{ij} = 1$ if $i \to j$ is an edge in $\mathcal{G}$). This choice of edge weights is sufficiently generic, since it results in a distribution that is faithful to $\mathcal{G}$. We then compute $\Lambda\Lambda^T$:

$$\Lambda = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad K = \Lambda\Lambda^T = \begin{pmatrix} 2 & 1 & -1 & 0 \\ 1 & 2 & -1 & 0 \\ -1 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Taking the log of the principal minors, we arrive at the system of inequalities:

$$\begin{array}{rrr} x_1 \leq \log 2 & x_1 + x_3 \leq \log 3 & x_1 + x_2 + x_3 \leq \log 4 \\ x_2 \leq \log 2 & x_1 + x_4 \leq \log 2 & x_1 + x_2 + x_4 \leq \log 3 \\ x_3 \leq \log 2 & x_2 + x_3 \leq \log 3 & x_1 + x_3 + x_4 \leq 0 \\ x_4 \leq 0 & x_2 + x_4 \leq \log 2 & x_2 + x_3 + x_4 \leq 0 \\ x_1 + x_2 \leq \log 3 & x_3 + x_4 \leq 0 & x_1 + x_2 + x_3 + x_4 = 0 \end{array}$$

For instance, the submatrix $K_{\{1,3\},\{1,3\}}$ is $\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, whose determinant is 3, giving the inequality $x_1 + x_3 \leq \log 3$. These inequalities give a realization of the DAG associahedron in Example 3.4. This realization is geometrically different from but has the same normal fan as the rational realization obtained by matroid polytopes that we will present in §5. □

As seen in the example above, there is a problem with this construction: The constant terms in the inequalities will almost never be all rational numbers, making it difficult to obtain exact combinatorial information such as the f-vector and the normal fan. We found that the following *heuristic* works well in practice to obtain exact combinatorial information from this polytope description: First, round off the real numbers in the inequalities to nearby rational numbers (e.g. using 52 bit precision). Then use exact arithmetic to compute the vertices of the polytope defined by these approximate inequalities. This results in approximations of the true vertices. Then form an approximate *slack matrix* by evaluating each approximate inequality at each approximate vertex and replace the entries in the slack matrix by 0 or 1 depending on whether the entry is approximately zero or not (e.g. by rounding off to 35 bit precision) to obtain an *incidence matrix* between the vertices and facets. By eliminating duplicate rows and columns from this matrix we obtain the incidence matrix of a new polytope, from which its face lattice can be computed. In our simulations, the incidence matrix obtained this way gives the correct number of vertices and facets of the DAG associahedron, at least in small dimensions. However, this does *not* immediately lead to a rational realization of the polytope. Our code is available on Github at `https://github.com/foxflo/DAG-associahedra`.

It is possible that in some (or even all) instances we may be able to choose the parameters ℓ_{ij}'s in such a way that the logarithms of the principal minors are all rational. However, we do not know of any systematic way to do this, nor do we know of a systematic way to transform a non-rational realization into a rational realization. We leave this as an open problem for future work. However, it is clear that if there is a non-rational realization, then there is also a rational realization, since a realization is a submodular function that satisfies some of the inequalities in (2) (those corresponding to the CI relations) at equalities and the rest as strict inequalities, and these linear constraints have rational coefficients.

To end this section, note that although for this paper it is sufficient to study the Gaussian setting, submodularity of the multiinformation holds for any probability distribution with finite multiinformation, which includes for example marginally continuous measures [Stu05]. Hence any set of CI relations that has a faithful realization by a distribution with finite multiinformation gives rise to a polytope similar to a DAG associahedron when contracting the edges correspond to CI relations in the permutohedron.

5. A construction of DAG associahedra as Minkowski sums of matroid polytopes

In the following, we obtain a construction of DAG associahedra as Minkowski sums of matroid polytopes, resulting in a rational realization of these polytopes. Until now we viewed a semigraphoid as defined by CI relations. However, we can equivalently define a semigraphoid by its complementary conditional dependence relations. Minkowski addition of generalized permutohedra translates to taking the union of the corresponding conditional dependence relations, since the union of normal cones to the edges of the Minkowski sum is the union of normal cones to the edges in the summand polytopes.[2]

[2] In other words, the tropical hypersurface of a Minkowski sum of polytopes is the union of the tropical hypersurfaces of individual polytopes.

For every dependence relation $i \not\perp\!\!\!\perp j \mid K$ in the semigraphoid defined by a DAG $\mathcal{G}$, we wish to find a matroid whose semigraphoid, defined by its rank function as in (4), contains the given relation and whose dependence relations are all valid for the semigraphoid of the DAG.

We now describe how to construct these matroids. For any conditional dependence relation $i \not\perp\!\!\!\perp j \mid K$ in the semigraphoid defined by a DAG $\mathcal{G}$, there is a Bayes ball path from i to j given K. We first partition the Bayes ball path into *canyons* and *treks* as follows.

Definition 5.1. A *trek* along a path is a consecutive subpath that does not contain any colliders. A *canyon* along a path is a consecutive subpath that is palindromic with exactly one collider in the middle such that all edges are directed toward the collider. A Bayes ball path is called *simple* if no node is repeated except in the same canyon and the maximal canyons are pairwise disjoint.

If we think of the arrows as always pointing down, then a canyon is a path that first goes down and then backtracks up the same edges to the first node. See Figure 5 for an example. A single collider is a canyon by itself but not necessarily a maximal one.

The *active paths* in [Sha98] can be obtained from simple Bayes ball paths by replacing each canyon with only the top of the canyon, e.g. for the Bayes ball path 1 4 8 4 3 (given $\{8\}$) we get an active path 1 4 3. We prefer to keep the canyons in the path because we will need them for our matroid construction below.

Lemma 5.2. *If there is a Bayes ball path from i to j given K, then there is a simple one that is an alternating sequence of disjoint treks and canyons, starting and ending with treks.*

Proof. Suppose there is a repeated node a. Then we can take the first edge into a and the last edge out of a. This is allowed except when a would become a collider on the new path and a is not in the conditioned set K. In this case there must be a descendant of a that is a collider, hence in K, so we can make a canyon between a and this collider. The same argument shows that we can make the maximal canyons to be pairwise disjoint and that the end nodes i and j are not in any canyons.

On the simple path, each connected component of the maximal canyons and their adjacent edges is a trek by definition, since it does not contain any colliders. Note that a single collider is considered a canyon. There must be at least one collider, hence a canyon, between any two such treks. For every canyon, we may assume that the node at the top must have two arrows pointing into it on the path; otherwise we can replace the canyon with just the top of the canyon to get another simple Bayes Ball path. If there are two consecutive canyons, then the edge between them cannot have an arrowhead at both canyons, so we can shortcut at least one of them. Thus we may assume that canyons do not occur next to each other, i.e. any two canyons are separated by a trek. ☐

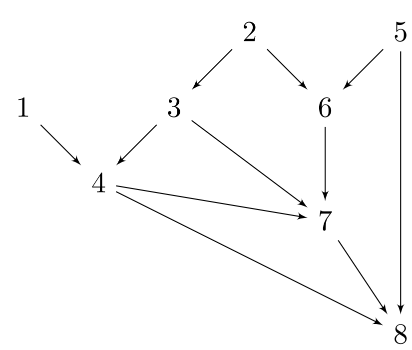


FIGURE 5. In the DAG, $\overline{1} \to \underline{4} \leftarrow \overline{3 \leftarrow 2} \to \underline{6 \to 7 \leftarrow 6} \leftarrow \overline{5 \to 8}$ is a Bayes ball path from 1 to 8 given $\{4, 7\}$. The treks and canyons along the path are overlined and underlined respectively.

For example, in Figure 5 the Bayes ball path 1 4 8 7 4 3 from 1 to 3 given $\{8\}$ has a repeated node 4, and simply removing the path between the two occurrences of 4 would give 1 4 3, which is *not* a Bayes ball path given $\{8\}$ since the collider 4 is not in the conditioned set $\{8\}$. However, 4 has a descendant, 8, which is a collider in the original Bayes Ball path, so we can create a canyon $4 \to 8 \leftarrow 4$ and take the path 1 4 8 4 3 instead.

Construction of a matroid from a simple Bayes ball path.

Let α be a simple Bayes ball path from i to j given K which is an alternating sequence of treks and canyons as in Lemma 5.2. Suppose we have $d+1$ treks $t_1, \ldots, t_{d+1}$ and d canyons $c_1, \ldots, c_d$, in the order $t_1 c_1 t_2 c_2 \cdots c_d t_{d+1}$. For $k = 1, \ldots, d$, let M_i be the rank 2 uniform matroid on $\{t_k, c_k, t_{k+1}\}$, i.e. a subset is independent if and only if it has size ≤ 2. It can be represented by affine independence among three distinct points on an affine line or as linear independence among three non-parallel vectors in a 2-dimensional vector space or as edges of a triangle.

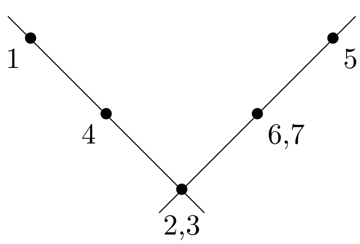
Let TC_α be the matroid on the set of treks and canyons $\{t_1, c_1, \ldots, t_d, c_d, t_{d+1}\}$, defined as the *parallel connection* or (free and proper) *amalgam* of these k matroids along the treks [Oxl11, §7, §11]. The parallel connection of two graphic matroids is obtained by gluing two graphs along an edge, which corresponds to a trek in our case. The parallel connection of two affine independence matroids is obtained by placing the affine spaces in a common ambient affine space in such a way that they only intersect at one point, which corresponds to a trek in our case. The matroid TC_α is constructed by repeating this operation, which is clearly associative.

Finally the matroid M_α on the node set $[n]$ of the DAG, is defined as follows. Let TC'_α be the matroid TC_α with an additional loop element ℓ. Let $f : [n] \to TC'_\alpha$ be a function that sends each element on the path α to the trek or canyon containing it and all other elements to the loop ℓ. We say that a subset $S \subset [n]$ is independent in the matroid M_α if $\{f(a) : a \in S\}$ is independent in TC'_α. In particular, elements in the same trek or the same canyon become parallel elements (two-element circuits). Two examples of such matroids for different Bayes ball paths are shown in Figure 6.

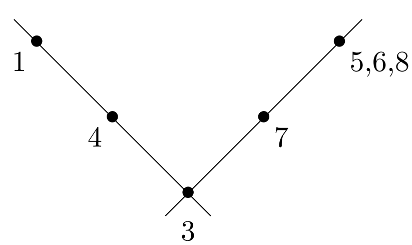
A subset S of a matroid is called a *flat* if $\operatorname{rank}(S \cup \{a\}) > \operatorname{rank}(S)$ for every $a \notin S$. The intersection of two flats is a flat. The *span* or the *closure* of a set is the smallest flat containing it. More precisely

$$\operatorname{span}(S) = \{a : \operatorname{rank}(S \cup \{a\}) = \operatorname{rank}(S)\}.$$

A subset $A \subseteq \{t_1, c_1, \ldots, t_d, c_d, t_{d+1}\}$ is a flat in TC_α if and only if $A \cap \{t_k, c_k, t_{k+1}\}$ is a flat for each $k = 1, \ldots, d$ [Oxl11, Proposition 11.4.13]. This can also be checked directly from the realization of TC_α using affine/linear independence or graphs. Note that a subset of $\{t_k, c_k, t_{k+1}\}$ is a flat if and only if it has size $\neq 2$. Flats of M_α are inverse images under f of flats in TC'_α.



(a) The matroid corresponding to the Bayes ball path $\overline{1}\,\underline{4}\,\overline{3\,2}\,\underline{6\,7\,6}\,\overline{5\,8}$, which goes from 1 to 8 given $\{4, 7\}$.



(b) The matroid corresponding to the Bayes ball path $\overline{1}\,\underline{4}\,\overline{3}\,\underline{7}\,\overline{6\,5\,8}$, which goes from 1 to 8 given $\{4, 7\}$. The element 2 is a loop in the matroid, i.e. $\{2\}$ is dependent.

FIGURE 6. Two matroids that are compatible with the DAG in Figure 5.

It follows that a subset $S \subset M_\alpha$ is a flat if and only if it satisfies all of the following conditions:

(F0) S contains all the loops (the nodes that are not on α)
(F1) If an element of a trek or a canyon is in S, then all the other nodes in the same trek or canyon are also in S.
(F2) For each $k = 1, \ldots, d$, if S intersects (thus contains) two out of three treks/canyons in $\{t_k, c_k, t_{k+1}\}$, then it also intersects (thus contains) the third.

□

Recall from §2 that the rank function of a matroid gives a collection of conditional dependence relations of the form $a \not\perp\!\!\!\perp b \mid C$, where $i, j \in [n]$ and $C \subset [n] \setminus \{i, j\}$ satisfy the condition (4), namely

$$\operatorname{rank}(C) + 1 = \operatorname{rank}(Ca) = \operatorname{rank}(Cab) = \operatorname{rank}(Cb) \tag{6}$$

Lemma 5.3. *Let $\mathcal{G}$ be a DAG and let α be a simple Bayes ball path from i to j given K in $\mathcal{G}$. Then the conditional dependence relations of the matroid M_α form a subset of the set of conditional dependence relations defined by the semigraphoid corresponding to $\mathcal{G}$.*

Proof. Suppose the relation $a \not\perp\!\!\!\perp b \mid C$ comes from the matroid M_α. We wish to show that there is a Bayes ball path in $\mathcal{G}$ between a and b given C. The condition (6) can be translated as

$$\operatorname{span}(C) \subsetneq \operatorname{span}(Ca) = \operatorname{span}(Cab) = \operatorname{span}(Cb). \tag{7}$$

Let us first consider the case when a and b are in the same trek or in the same canyon. Then C cannot contain any element from the same trek/canyon; otherwise both a and b would be in $\operatorname{span}(C)$, contradicting (7). Any two nodes in the same trek or the same canyon are connected by a Bayes ball path if no node is conditioned. Thus there is a Bayes ball path between a and b given C, along α.

Now suppose that a and b are in different treks/canyons. Then condition (7) implies that $\operatorname{span}(C)$, hence C, cannot contain any element in the treks/canyons containing a or b.

We claim that C does not intersect any trek that lies strictly between a and b along α. Otherwise, if we compute $\operatorname{span}(Ca)$ by adding to $\operatorname{span}(C)$ nodes along the path starting at a, then the process would terminate (i.e. the conditions (F0),(F1),(F2) would be satisfied) at or before the trek that intersects C, before it reaches b. Thus $b \notin \operatorname{span}(Ca)$, contradicting the condition $\operatorname{span}(Ca) = \operatorname{span}(Cb)$ in (7).

Next we claim that C intersects every canyon that lies strictly between a and b along α. Suppose C does not intersect a canyon. But we have already shown that C does not intersect the next trek (which may contain b) after the canyon, on the path from a to b along α. Then as before, the computation of $\operatorname{span}(Ca)$ stops before it reaches b; so again $b \notin \operatorname{span}(Ca)$.

Putting everything together we conclude that C intersects all the canyons and none of the treks that lie strictly between a and b along α. This means that the subpath between a and b along α forms a Bayes ball path given C in $\mathcal{G}$, and the relation $a \not\perp\!\!\!\perp b \mid C$ is valid for the semigraphoid of $\mathcal{G}$. □

A proof of our main result now follows.

Proof of Theorem 3.2. For every conditional dependence relation $i \not\perp\!\!\!\perp j \mid K$ in the semigraphoid corresponding to a DAG $\mathcal{G}$, we find a simple Bayes ball path α from i to j given K and construct a matroid M_α. By Lemma 5.3, all the conditional dependence relations coming from M_α are among those in the semigraphoid corresponding to $\mathcal{G}$. By taking all the matroids of the form M_α where α runs over all simple Bayes ball paths in $\mathcal{G}$, we obtain exactly all the conditional dependence relations that are valid in the semigraphoid of $\mathcal{G}$. Taking the union of all these dependence relations

translates into taking the Minkowski sum of all the corresponding matroid polytopes. Hence, DAG associahedra are obtained as Minkowski sums of matroid polytopes. □

Example 5.4 (Example 3.4 continued)**.** Consider the DAG with 4 nodes from Example 3.4. The path $\overline{1}\,\underline{3\,4\,3}\,\overline{2}$ is a Bayes ball path from 1 to 2 given $\{4\}$. The corresponding rank 2 matroid is realized by affine dependence among 3 distinct points in $\mathbb{R}$. For example, we can take the matroid on the points $v_1 = 1, v_2 = 2, v_3 = v_4 = 3$ on the real line $\mathbb{R}$ under affine independence. Equivalently, the matroid is given by columns of the matrix $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 3 \end{pmatrix}$ under linear independence. The bases are all pairs $\{v_i, v_j\}$ where $i \neq j$, except the pair $\{v_3, v_4\}$. The matroid polytope is a square-based pyramid, with $e_1 + e_2$ as the tip of the pyramid. □

6. Generalizing to mixed graphs

The MSMP construction generalizes to a much more general setting of semigraphoids arising from loopless mixed graphs (LMG) introduced by Sadeghi and Lauritzen in [SL14]. We first recall the definitions. A *mixed graph* is a graph with three possible types of edges: undirected (i—j), directed ($i \longleftarrow j$ or $i \longrightarrow j$), or bidirected ($i \longleftrightarrow j$), which are also called *lines*, *arrows*, and *arcs* respectively. Multiple types of edges are allowed between any two nodes. A loopless mixed graph is a mixed graph without a loop, or an edge between a node and itself.

A node j is called an *ancestor* of a node i, and i is called a descendent of j, if there is a path $i = i_0, i_1, \ldots, i_n = j$ from i to j in which the edges (i_k, i_{k+1}) are arrows (directed edges) pointing from i_k to i_{k+1} for all $k = 0, \ldots, n-1$. Note that undirected and bidirected edges are not used in the definition of ancestors. The set of ancestors of a node i is denoted by $\mathrm{an}(i)$. For any set of nodes K, let $\mathrm{an}(K) = \bigcup_{k \in K} \mathrm{an}(k)$.

For the path ijk (where we may have $i = k$) the node j is called a *collider* if the path is one of $i \longrightarrow j \longleftarrow k$, $i \longleftrightarrow j \longleftarrow k$, or $i \longrightarrow j \longleftrightarrow k$. Otherwise k is a non-collider.

Let K be a subset of the node set of an LMG. A path is called a *Bayes ball path given* K (called an *m-connecting path given* K in [SL14]) if all its collider nodes are in $K \cup \mathrm{an}(K)$ and all its non-collider nodes are outside K. We say that $i \not\perp\!\!\!\perp j \mid K$ if there exists a Bayes ball path from i to j given K. This collection of CI relations forms a graphoid [SL14].

We define *treks*, *canyons*, and *simple* Bayes ball paths in an LMG in exactly the same way as in Definition 5.1. Only directed edges (arcs) can appear in a canyon, but all three types of edges are allowed on a trek. A trek or a canyon may consist of only one node, but it may not be empty.

Lemma 6.1. *Let i, j be nodes in an LMG and K be a set of nodes such that $\{i, j\} \cap K = \varnothing$. If there is a Bayes ball path from i to j given K, then there is a simple one that is a sequence of treks and canyons, starting and ending with treks, such that*

1. *between any two treks there is at least one canyon*
2. *on the edge between a trek and a canyon, there must be an arrowhead at the canyon, and*
3. *two consecutive canyons can only be connected by a bidirected edge.*

Proof. The existence of a simple path follows from the same argument as in the proof of Lemma 5.2. On the simple path, each connected component of the complement of maximal canyons (and adjacent edges) is a trek, since it does not contain any collider. Note that a single collider is considered a canyon. The property (1) follows from the construction of treks as connected components. Consider an edge connecting a trek and a canyon. If there is no arrowhead at the canyon, then the top of the canyon does not have two arrowheads pointing into it. We can then shortcut the canyon, replacing

the canyon with only the top of it, to get another simple Bayes ball path. Thus property (2) is satisfied. An anologous argument shows that an edge between two canyons must have arrowheads at both canyons; otherwise we can shortcut the canyons. Thus property (3) is also satisfied. □

Now we can describe a generalization of the matriod construction from the previous section.

Construction of a matroid from a Bayes ball path in an LMG. Let α be a simple Bayes ball path from i to j given K satisfying the conditions from Lemma 6.1. Suppose there are $d+1$ treks $t_1, \ldots, t_{d+1}$ on α, in this order. For $k = 1, \ldots, d$, let m_k denote the number of canyons between t_k and t_{k+1}. For each subpath $t_k c_{k,1} \cdots c_{k,m_k} t_{k+1}$ of two treks separated by canyons, consider the uniform matroid U_{m_k+1, m_k+2} on $\{t_k, c_{k,1}, \ldots, c_{k,m_k}, t_{k+1}\}$, which can be represented by affine independence among $m_k + 2$ general points in $\mathbb{R}^{m_k}$. We then take the parallel connection of these uniform matroids along the treks. In other words, we place the affine spaces, one for each pair of treks separated by a sequence of canyons, in a common ambient space so that any two consecutive ones only meet at one point and they affinely span maximum possible dimension. As before, a subset is a flat if and only if its intersection with each of the uniform matroids is also a flat.

The matroid M_α on the node set $[n]$ of the LMG is defined as before by replacing each trek (resp. canyon) with parallel elements corresponding to the nodes in the trek (resp. canyon) and considering nodes not on α as loops.

A subset S of the node set $[n]$ is a flat in M_α if and only if it satisfies (F0), (F1), and

(F2′) For each $k = 1, \ldots, d$, if S intersects (thus contains) $m_k + 1$ out of $m_k + 2$ treks/canyons in $\{t_k, c_{k,1}, \ldots, c_{k,m_k}, t_{k+1}\}$, then it also intersects (thus contains) the remaining one.

□

For example, for the Bayes ball path $t_1 \longleftrightarrow c_{1,1} \longleftrightarrow c_{1,2} \longleftarrow t_2 \longleftrightarrow c_{2,1} \longleftrightarrow c_{2,2} \longleftarrow t_3$, we can take the following representation via affine independence in $\mathbb{R}^4$:

trek or canyon	t_1	$c_{1,1}$	$c_{1,2}$	t_2	$c_{2,1}$	$c_{2,2}$	t_3
	0	1	2	3	3	3	3
representation	0	1	4	9	9	9	9
as points in $\mathbb{R}^4$	0	0	0	0	1	2	3
	0	0	0	0	1	4	9

We have a rank 3 uniform matroid on the first four elements and another rank 3 uniform matroid on the last four elements, meeting at a point t_2.

Theorem 6.2 (Generalization of Main Theorem)**.** *The semigraphoid of any loopless mixed graph is submodular. The associated coarsening of the S_n-fan is the normal fan of the Minkowski sum of matroid polytopes corresponding to simple Bayes ball paths satisfying the conditions in Lemma 6.1.*

Proof. The result and the proof of Lemma 5.3 and the Proof of Theorem 3.2 in §5 can be repeated word for word, with the word DAG replaced by LMG and the condition (F2) replaced by (F2′). □

7. Relationship among families of semigraphoids

Recall from Lemma 2.3 that submodular functions on $2^{[n]}$ correspond to polytopal coarsenings of the S_n fan, and these are exactly the normal fans of generalized permutohedra.

Graph associahedra and DAG associahedra are special classes of generalized permutohedra that are defined up to equivalence of normal fans. The former can be realized as Minkowski sums of standard simplices and the latter can be realized as Minkowski sums of matroid polytopes (MSMP). What additional classes of generalized permutohedra can be realized in this way? Since the standard simplices are matroid polytopes, MSS polytopes are also MSMP.

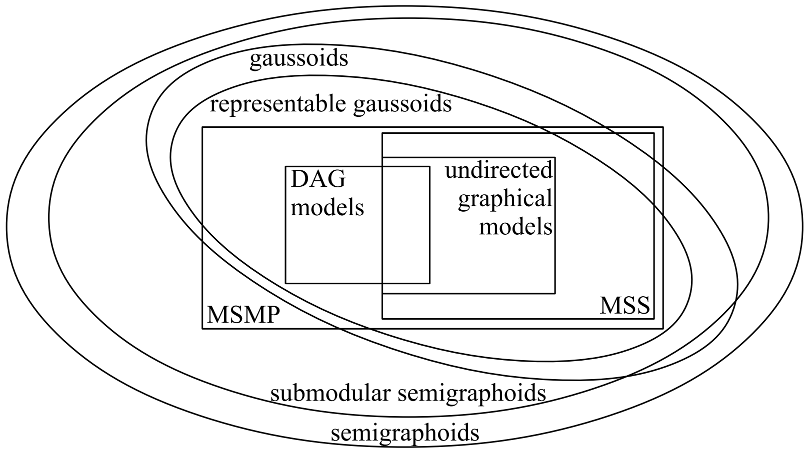


FIGURE 7. Venn diagram representing the relationship of all the different coarsenings of the S_n fan discussed in this paper.

Unfortunately, this question seems difficult to answer in general. For $n = 3, 4, 5$ respectively, the cone of submodular functions has 5, 37, and 117978 extreme rays of which only 5, 23, and 149 respectively correspond to (connected) matroid polytopes. It suffices to consider connected matroids because the direct sum of matroids corresponds to the Minkowski sum of the corresponding matroid polytopes. Thus the matroid polytope of a disconnected matroid, which is the direct sum of non-trivial matroids, is the Minkowski sum of the matroid polytopes of these direct summands. Although the structure of these extreme rays is unclear, it seems unlikely due to their sparsity that many submodular semigraphoids will arise in this way.

Another interesting class of semigraphoids are gaussoids [LM07], an abstraction of regular Gaussian distributions in the language of CI relations; see §2. Since we have seen that probabilistic graphical models can be faithfully realized by regular Gaussian distributions, another natural question is whether all regular Gaussian models (also called representable gaussoids) or even all gaussoids are MSMP. Our interest in gaussoids stems from Theorem 8.3, where gaussoids give a natural setting.

Gaussoids appear to be incompatible with the MSMP construction. We have computationally verified that for $3 \leq n \leq 8$ no submodular semigraphoid corresponding to a connected matroid on $[n]$ is a gaussoid. Thus, none of the extreme matroidal rays of the submodular cone are gaussoids.

Conversely, not all gaussoids, in fact not even all representable gaussoids, can be obtained via MSMP. For example, [DX10, table A.1] lists all Gaussian CI models on four variables (up to equivalence) and examples 19, 20, 34, 50, 51 are not MSMP. On the other hand, the CI relations corresponding to graphical models in this list all correspond to generalized permutohedra arising as MSMP.

In Figure 7 we illustrate the relationship of all the different coarsenings of the S_n fan discussed in this paper by a Venn diagram. We have seen that undirected graphical models give rise to MSSs, while DAG models can be realized by MSMPs. In Proposition 3.6 we showed that a DAG model is MSS if and only if it coincides with an undirected graphical model, i.e. if and if it is a decomposable model. As we have discussed above, gaussoids are incompatible with the MSMP construction. In fact, gaussoids are also incompatible with the MSS construction. For example, it is easy to check that the standard simplex in Figure 2(a) is not a gaussoid. While every representable gaussoid is a submodular gaussoid as shown in Lemma 4.1, this is not the case for gaussoids. The semigraphoid studied in [HMS$^+$08, Section 3] is a gaussoid that is not submodular.

8. Causal inference

In this section, we describe how DAG associahedra can be used to perform causal inference. The main problem in causal inference is the following: We obtain data from an unobserved DAG $\mathcal{G}$. From this data we infer a set of CI relations $\mathcal{C}$. Under the faithfulness assumption, which we will assume throughout this section, $\mathcal{C}$ coincides with the gaussoid of $\mathcal{G}$. The goal is to learn $\mathcal{G}$ from $\mathcal{C}$. This problem is ill-defined since d-separation does not uniquely identify a DAG. So instead the problem is to learn $\mathcal{G}$ up to Markov equivalence, or in other words, to learn from $\mathcal{C}$ the *essential graph*, which is a partially directed graph with the same skeleton as $\mathcal{G}$ where an edge is directed if and only if it is directed the same way in every DAG in the Markov equivalence class.

A popular algorithm for learning the Markov equivalence class of a DAG is Greedy Equivalence Search (GES) [Mee97, Chi02b], a greedy algorithm that searches through the space of DAGs by maximizing a scoring criterion such as the Bayesian Information Criterion (BIC). Under the faithfulness assumption GES is known to be consistent, i.e. it learns the correct essential graph with probability approaching 1 as the sample size goes to infinity [Mee97, Chi02b]. To reduce computation time, Teyssier and Koller [TK05] suggested to replace the greedy search in DAG space by a greedy search in the space of all orderings; a scoring criterion such as BIC is optimized by performing a walk on the edges of the permutohedron. Although no consistency guarantees were given for this greedy algorithm, simulations suggest that the greedy ordering-based search has a similar performance and lower computational costs as compared to GES [TK05]. In the following, we use our geometric insight on DAG associahedra to develop a new greedy ordering-based search with consistency guarantees.

Let F be a coarsening of the S_n fan. Each cone in F is defined by inequalities of the form $x_i \leq x_j$ and can be labeled a poset on $[n]$. Then we get a map from permutations of $[n]$ to the set of partial orders on $[n]$, derived from the map sending a maximal S_n cone to the maximal cone F containing it. The preimage permutation (total order) is a linear extension of its image partial order. Hence the maximal cones of the coarsened S_n fan — or the vertices of the generalized permutohedron if the fan is polytopal — can be labeled by posets so that every permutation is a linear extension of exactly one of the posets. If two permutations π and τ are mapped to the same partial order, then we denote this by $\pi \sim \tau$.

A semigraphoid $\mathcal{C}$ on $[n]$ also gives a map from S_n to the set of DAGs on nodes $[n]$ as described in [RU14]: To every permutation π we associate a DAG $\mathcal{G}_\pi$ with

$$(\pi_i, \pi_j) \in \mathcal{G}_\pi \iff i < j \text{ and } \pi_i \not\perp\!\!\!\perp \pi_j \mid \{\pi_1, \dots, \pi_{\max(i,j)}\} \setminus \{\pi_i, \pi_j\}. \tag{8}$$

In other words, the edge directions in the graph must be compatible with the ordering $\pi = (\pi_1|\pi_2|\cdots|\pi_n)$, and the existence of an edge means that the two nodes are *not* independent given all the nodes that come before them in the ordering. G_π is also known as a minimal I-map or a directed independence graph.

We call π a topological ordering of $\mathcal{G}$ if any edge (i,j) in $\mathcal{G}$ implies that $i \succ j$ in π. Note that if the semigraphoid comes from a DAG $\mathcal{G}$ and π is a topological ordering of $\mathcal{G}$, then $\mathcal{G} = \mathcal{G}_\pi$.

In [RU14], it was proposed to use the number of edges of $\mathcal{G}_\pi$ as a scoring criterion. It was shown that an algorithm that outputs the Markov equivalence class of $\mathcal{G}_\pi$ with the fewest number of edges is consistent, i.e. it outputs the correct Markov equivalence class, under strictly weaker conditions than faithfulness. A permutation π giving a sparsest DAG is called a *sparsest permutation.* However the sparsest permutation (SP) algorithm is problematic from a computational point of view since it requires searching over all permutations. Instead, similarly as suggested in [TK05], we can perform a greedy search by traversing the edges of the permutohedron, using the number of edges of $\mathcal{G}_\pi$ as a scoring function (see Algorithm 1).

Algorithm 1 Greedy SP algorithm on the permutohedron

Input: A set of CI relations $\mathcal{C}$ on n random variables and a starting permutation $\pi \in S_n$

Output: An essential graph G.

(1) Set $t := 0$ and $\pi^{(0)} := \pi$.
(2) Set $t := t+1$. Randomly select a permutation $\pi^{(t)}$ that differs from $\pi^{(t-1)}$ in a single adjacent transposition such that $\mathcal{G}_{\pi^{(t)}}$ is at least as sparse as $\mathcal{G}_{\pi^{(t-1)}}$.
(3) Iterate (2) until convergence to the sparsest Markov equivalence class and output the corresponding essential graph.

Algorithm 1 requires searching through neighboring permutations even when they give rise to the same DAG. For example, the neighboring permutations $\pi = (1|2|3|4)$ and $\tau = (2|1|3|4)$ in Example 3.4 give rise to the same DAG $\mathcal{G}_\pi = \mathcal{G}_\tau = \mathcal{G}$ shown in Figure 3 (left). We next discuss how to reduce the search space and hence computation time by performing the greedy search on the smaller DAG associahedron instead of the full permutohedron. The difficulty is that this needs to be done without having access to the DAG $\mathcal{G}$ on which the DAG associahedron is based. In order to do this, we give a description of the vertices and edges of a DAG associahedron in terms of the DAGs $\mathcal{G}_\pi$ that are associated to its vertices.

Theorem 8.1. *For any fixed graphoid and two permutations π and τ, we have*

$$\pi \sim \tau \iff \mathcal{G}_\pi = \mathcal{G}_\tau.$$

Moreover, the equivalence class of π consists of all topological orderings of $\mathcal{G}_\pi$.

Proof. Suppose $\pi \sim \tau$. We may assume that

$$\pi = (a_1|\cdots|a_k|i|j|b_1|\dots|b_{n-k-2}) \quad \text{and} \quad \tau = (a_1|\cdots|a_k|j|i|b_1|\dots|b_{n-k-2}),$$

where $i \perp\!\!\!\perp j \mid \{a_1,\dots,a_k\}$, since any pair of equivalent permutations is connected by a sequence of such pairs. Now let us compare the edges in $\mathcal{G}_\pi$ and $\mathcal{G}_\tau$. There is no edge between i and j in either DAG. Between any two nodes in $[n]\backslash\{i,j\}$, it is clear that $\mathcal{G}_\pi$ and $\mathcal{G}_\tau$ coincide.

Now suppose that (a_ℓ, j) is not an edge in $\mathcal{G}_\pi$ for some ℓ. Let $K = \{a_1,\dots,a_k\}\backslash\{a_\ell\}$. Then by applying the intersection property (INT) of graphoids from §2 we obtain

$$i \perp\!\!\!\perp j \mid K a_\ell \text{ and } j \perp\!\!\!\perp a_\ell \mid Ki \overset{\text{(INT)}}{\Longrightarrow} j \perp\!\!\!\perp a_\ell \mid K.$$

Thus (a_ℓ, j) is not an edge in $\mathcal{G}_\tau$ either. Similarly if (a_ℓ, j) is not an edge in $\mathcal{G}_\tau$, then applying the semigraphoid property (SG2) we obtain

$$j \perp\!\!\!\perp a_\ell \mid K \text{ and } i \perp\!\!\!\perp j \mid K \cup \{a_\ell\} \overset{\text{(SG2)}}{\Longrightarrow} j \perp\!\!\!\perp a_\ell \mid Ki.$$

Thus (a_ℓ, j) is not an edge in $\mathcal{G}_\pi$ either.

We can check in a similar fashion by setting $K = \{a_1,\dots,a_k\}$ that for any $b_\ell \in [n]\backslash(\{a_1,\dots,a_k\} \cup \{i,j\})$ the edge (j, b_ℓ) is in $\mathcal{G}_\pi$ if and only if it is in $\mathcal{G}_\tau$. The same claims also hold for i by switching π and τ.

For the converse, suppose τ is a topological ordering of $\mathcal{G}_\pi$. In particular, this holds when $\mathcal{G}_\pi = \mathcal{G}_\tau$. We wish to prove that $\pi \sim \tau$. Without loss of generality we may assume that $\tau = (1|2|\cdots|n)$. Let $\pi = (\pi_1|\pi_2|\cdots|\pi_n)$. If $\pi \neq \tau$, then there is an $i \in [n-1]$ such that $\pi_i > \pi_{i+1}$. Since π_i and π_{i+1} appear with opposite orders in π and τ and τ is a topological ordering of $\mathcal{G}_\pi$, there is no edge between π_i and π_{i+1} in $\mathcal{G}_\pi$. By construction of $\mathcal{G}_\pi$, we must have $\pi_i \perp\!\!\!\perp \pi_{i+1} \mid \{\pi_1,\dots,\pi_{i-1}\}$ in the graphoid. Let $\pi' = (\pi_1|\cdots|\pi_{i-1}|\pi_{i+1}|\pi_i|\pi_{i+2}|\cdots|\pi_n)$. Then $\pi' \sim \pi$ by definition, so $\mathcal{G}_{\pi'} = \mathcal{G}_\pi$ as shown above. Since τ is also a topological ordering of $\mathcal{G}_{\pi'}$, the statement $\tau \sim \pi$ follows by induction on the number of inversions in π. $\square$

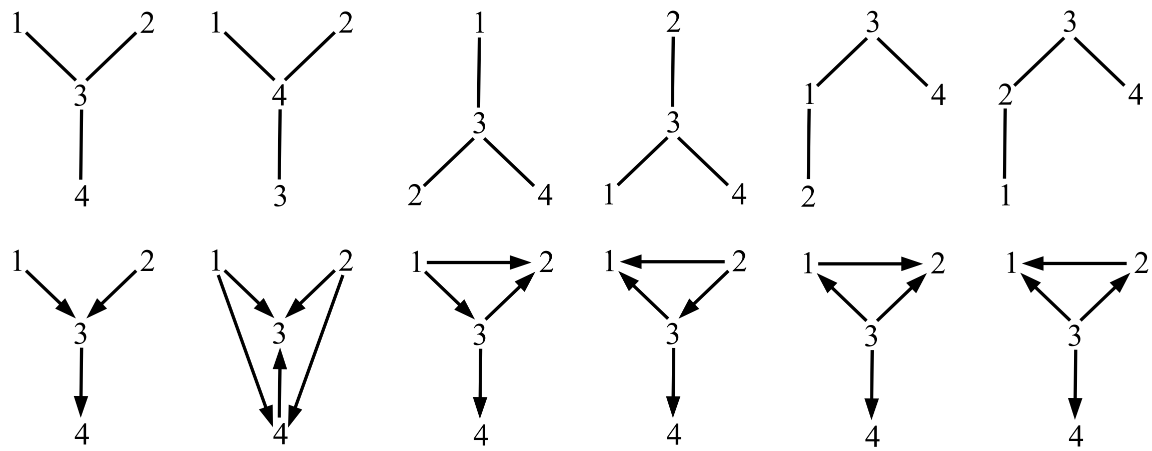


FIGURE 8. Posets and their corresponding DAGs representing the new (compared to the permutohedron) vertices of the DAG associahedron discussed in Examples 3.4 and 8.2.

In the following example we illustrate Theorem 8.1 and show how the vertices of a DAG associahedron can be labeled by posets or by DAGs.

Example 8.2. We return to Example 3.4. Compared to the permutohedron, the DAG associahedron corresponding to $\mathcal{G}$ has six new vertices, namely:

(a) $(1|2|3|4), (2|1|3|4)$,
(b) $(1|2|4|3), (2|1|4|3)$,
(c) $(1|3|2|4), (1|3|4|2)$,
(d) $(2|3|1|4), (2|3|4|1)$,
(e) $(3|4|1|2), (3|1|4|2), (3|1|2|4)$,
(f) $(3|4|2|1), (3|2|4|1), (3|2|1|4)$.

The posets representing these vertices and the corresponding DAGs are shown in Figure 8. Each of the other vertices of the DAG associahedron corresponds to a single permutation and the corresponding DAG has no missing edges. □

If we have a description of the vertices of a DAG associahedron in terms of posets, then we know the maximal cones in the normal fan, so we can directly obtain all other normal cones by intersecting the maximal cones. In the following, we give an alternative description of the edges of the DAG associahedron in terms of the DAGs $\mathcal{G}_\pi, \mathcal{G}_\tau$ corresponding to the vertices adjacent to an edge (π, τ).

Chickering [Chi95] introduced the notion of a covered edge: a directed edge (i, j) in $\mathcal{G}$ is *covered* if

$$\mathrm{pa}(i) = \mathrm{pa}(j) \setminus \{i\}.$$

We denote by $\overline{\mathcal{G}}$ the skeleton of a DAG $\mathcal{G}$. In addition, for two undirected graphs G and G' we say that G is a subset of G', i.e., $G \subseteq G'$, if G and G' have the same node set and every edge in G is also an edge in G'.

The following result shows that given a DAG label of a vertex of a DAG associahedron, we can find neighboring vertices whose underlying graph is not bigger by flipping the direction of a covered edge. We will prove this result more generally for gaussoids.

Theorem 8.3. *Let F be a coarsened S_n fan corresponding to a gaussoid. Suppose the equivalence classes of $\pi = (\pi_1|\pi_2|\cdots|\pi_n)$ and $\tau = (\pi_1|\pi_2|\cdots|\pi_{i+1}|\pi_i|\cdots|\pi_n)$ are adjacent maximal cones in F. Then $\overline{\mathcal{G}}_\tau \subseteq \overline{\mathcal{G}}_\pi$ if and only if (π_i, π_{i+1}) is a covered edge in $\mathcal{G}_\pi$.*

Proof. First, note that (π_i, π_{i+1}) is an edge in $\mathcal{G}_\pi$, since otherwise $\mathcal{G}_\pi = \mathcal{G}_\tau$ by Theorem 8.1. We now prove the "if" direction. Without loss of generality we assume that $\pi = (1|2|\cdots|n)$, $\tau = (1|2|\cdots|i-1|i+1|i|i+2|\cdots|n)$ and $(i, i+1)$ is a covered edge in $\mathcal{G}_\pi$. Note that from the definition of $\mathcal{G}_\pi$ and $\mathcal{G}_\tau$ the only difference between these two DAGs can be in the presence or absence of edges (ℓ, i) or $(\ell, i+1)$ with $\ell < i$. In order to prove that $\overline{\mathcal{G}_\tau} \subseteq \overline{\mathcal{G}_\pi}$, we need to show that any missing edge (ℓ, i) or $(\ell, i+1)$ in $\mathcal{G}_\pi$ is also not present in $\mathcal{G}_\tau$. Now suppose that (ℓ, i) is a missing edge in $\mathcal{G}_\pi$ for some $\ell < i$. Since the edge $(i, i+1)$ is covered in $\mathcal{G}_\pi$, then $(\ell, i+1)$ is also a missing edge in $\mathcal{G}_\pi$. Let $K = \{1, \ldots, i-1\} \backslash \{\ell\}$. By the definition of $\mathcal{G}_\pi$ and $\mathcal{G}_\tau$ we get that

$$\ell \perp\!\!\!\perp i \mid K \quad \text{and} \quad \ell \perp\!\!\!\perp i+1 \mid Ki,$$

and hence by the semigraphoid property (SG2) we obtain that $\ell \perp\!\!\!\perp i+1 \mid K$ and $\ell \perp\!\!\!\perp i \mid K \cup \{i+1\}$. Therefore, (ℓ, i) and $(\ell, i+1)$ are also missing edges in $\mathcal{G}_\tau$, and we conclude that $\overline{\mathcal{G}_\tau} \subseteq \overline{\mathcal{G}_\pi}$.

For the "only if" direction suppose that $\overline{\mathcal{G}_\tau} \subseteq \overline{\mathcal{G}_\pi}$. We want to show that the edge (π_i, π_{i+1}) is a covered edge in $\mathcal{G}_\pi$. Assume on the contrary that it is not.

We first consider the case when there is an $a < i$ with $(a, i+1) \in \mathcal{G}_\pi$ but $(a, i) \notin \mathcal{G}_\pi$. Then $(a, i) \notin \mathcal{G}_\tau$, and hence

$$a \perp\!\!\!\perp i \mid K \cup \{i+1\}, \tag{9}$$

where $K = \{1, \ldots, i-1\} \backslash \{a\}$. We claim that $(a, i+1) \in \mathcal{G}_\tau$. Otherwise we would have $a \perp\!\!\!\perp i+1 \mid K$, which together with (9) implies $a \perp\!\!\!\perp i+1 \mid Ki$ by (SG2), contradicting $(a, i+1) \in \mathcal{G}_\pi$. From $(a, i+1) \in \mathcal{G}_\tau$, we have

$$a \not\perp\!\!\!\perp i+1 \mid K. \tag{10}$$

Next we claim that

$$i \not\perp\!\!\!\perp i+1 \mid K. \tag{11}$$

Otherwise, together with (9) we would have $i \perp\!\!\!\perp i+1 \mid Ka$ by (SG2), contradicting the assumption that π and τ lie in adjacent cones of the fan F. Finally, from the weak-transitivity axiom (G2) for gaussoids we obtain

$$i \not\perp\!\!\!\perp i+1 \mid K \quad \text{and} \quad a \not\perp\!\!\!\perp i+1 \mid K \overset{\text{(G2)}}{\Longrightarrow} a \not\perp\!\!\!\perp i \mid K \quad \text{or} \quad a \not\perp\!\!\!\perp i \mid K \cup \{i+1\} \tag{12}$$

Combining (9) and (12), we obtain $a \not\perp\!\!\!\perp i \mid K$, that is, $(a, i) \in \mathcal{G}_\pi$, contradicting the assumption that $(a, i) \notin \mathcal{G}_\pi$.

Now we consider the case where $(a, i) \in \mathcal{G}_\pi$, but $(a, i+1) \notin \mathcal{G}_\pi$, so $(a, i+1) \notin \mathcal{G}_\tau$. Then

$$a \perp\!\!\!\perp i+1 \mid Ki \quad \text{and} \quad a \perp\!\!\!\perp i+1 \mid K. \tag{13}$$

By the gaussoid axiom (G1), we have

$$i+1 \not\perp\!\!\!\perp a \mid Ki \quad \text{or} \quad i+1 \not\perp\!\!\!\perp i \mid Ka \;\Rightarrow\; a \not\perp\!\!\!\perp i+1 \mid K \quad \text{or} \quad i \not\perp\!\!\!\perp i+1 \mid K. \tag{14}$$

Since π and τ are in adjacent cones of the fan F, we have $i+1 \not\perp\!\!\!\perp i \mid Ka$, so by (13) and (14),

$$i \not\perp\!\!\!\perp i+1 \mid K. \tag{15}$$

Since by assumption $(a, i) \in \mathcal{G}_\pi$, then $a \not\perp\!\!\!\perp i \mid K$. This together with (15) and weak transitivity (G2) gives us $a \not\perp\!\!\!\perp i+1 \mid Ki$ or $a \not\perp\!\!\!\perp i+1 \mid K$, which contradicts (13). $\square$

Algorithm 2 Greedy SP algorithm on the DAG associahedron

Input: A set of CI relations $\mathcal{C}$ on n random variables and a starting permutation $\pi \in S_n$

Output: An essential graph G.

(1) Set $t = 0$ and $\pi^{(0)} = \pi$.
(2) Set $t := t + 1$. Randomly select a covered edge $(\pi_i^{(t-1)}, \pi_j^{(t-1)})$ in $\mathcal{G}_{\pi^{(t-1)}}$ and reverse its direction. Let $\pi^{(t)}$ denote the resulting permutation and $\mathcal{G}_{\pi^{(t)}}$ the corresponding DAG.
(3) Iterate (2) until convergence to the sparsest Markov equivalence class and output the corresponding essential graph.

This result directly gives rise to an improved version of Algorithm 1, which corresponds to performing a greedy search on the DAG associahedron instead of the permutohedron and does not require knowing the underlying true DAG (see Algorithm 2). In this algorithm, we are given a set of CI relations $\mathcal{C}$ that are induced from a fixed but unknown DAG $\mathcal{G}$. In each iteration the algorithm outputs an auxiliary DAG, whose skeleton contains the skeleton of $\mathcal{G}$. We end by providing a sketch of the proof that this algorithm converges to $\mathcal{G}$ under the faithfulness assumption. The complete proof can be found in [MSUW17], a statistical follow-up work, where we show the importance of the geometric results obtained in this paper for applications to causal inference.

Theorem 8.4. *Algorithm 2 is consistent under the faithfulness condition.*

Proof. Let $\mathcal{G}$ denote the true DAG. Then $\mathcal{G} = \mathcal{G}_\pi$ for some π (any topological ordering of $\mathcal{G}$). Let $\tau \in S_n$. Then every independence relation that holds for $\mathcal{G}_\tau$ also holds for $\mathcal{G}$ [RU14, Lemma 2.1]. This implies $\overline{\mathcal{G}} \subseteq \overline{\mathcal{G}_\tau}$. If a permutation π differs from τ only in the reversal of a covered edge in $\mathcal{G}_\tau$, then by Theorem 8.3 we have $\overline{\mathcal{G}_\pi} \subseteq \overline{\mathcal{G}_\tau}$. At a high level, the proof follows from a result by Chickering [Chi02b, Theorem 4] which says that using such edge reversals one can go from any DAG $\mathcal{G}_\tau$ to any DAG $\mathcal{G}_\pi$ with $\overline{\mathcal{G}_\pi} \subseteq \overline{\mathcal{G}_\tau}$. The difficulty lies in showing that there exists such a Chickering sequence which corresponds to a walk on the DAG associahedron. This is proven in [MSUW17, Theorem 17]. □

Appendix A. Polytopes and fans

Most of the following definitions can be found in [Zie95]. A *polytope* is the convex hull of a finite set of points in a real vector space $\mathbb{R}^d$. The *Minkowski Sum* of two polytopes P and Q is defined as $P + Q = \{x + y \mid x \in P, y \in Q\}$. A (polyhedral) *cone* is the set of all non-negative linear combinations of a finite set of vectors in $\mathbb{R}^n$. A *face* of a polytope or a cone P is a subset of P that maximizes some linear functional. A facet is an inclusion-maximal face.

A *fan* is a family $\mathcal{F}$ of non-empty polyhedral cones such that

(1) Every non-empty face of a cone in $\mathcal{F}$ is also a cone in $\mathcal{F}$.
(2) The intersection of any two cones in $\mathcal{F}$ is a face of both.

A fan in $\mathbb{R}^n$ is *complete* if the union of its cones is equal to $\mathbb{R}^n$. A *wall* in a complete fan in $\mathbb{R}^n$ is an $d-1$-dimensional cone in the fan.

For each non-empty face F of P, the *outer normal cone* N_F is the set of all linear functionals that are maximized on F, i.e.

$$N_F = \{c \in (\mathbb{R}^d)^* \mid F \subset \{x \in P \mid c \cdot x = \max_{y \in P}(c \cdot y)\}\}.$$

The outer *normal fan* of a polytope P is the collection $\{N_F : F \text{ is a non-empty face of } P\}$, which is a complete fan in $\mathbb{R}^n$. The *inner normal* cones and fans are defined analogously by replacing

"max" with "min". For two faces F and F', if $F \subset F'$, then $N_F \supset N_{F'}$. In particular, the rays of the normal fan are the facet normal vectors of P, and the full dimensional cones in the normal fans are normal cones of the vertices of P.

Appendix B. A proof of Lemma 2.3

Lemma 2.3. A polytope $P \subset \mathbb{R}^n$ is a generalized permutohedra if and only if there exists a submodular function $\omega : 2^{[n]} \to \mathbb{R}$ with $\omega(\varnothing) = 0$ such that

$$P = \{x \in \mathbb{R}^n : \sum_{i \in I} x_i \leq \omega(I) \text{ for each non-empty } I \subset [n], \text{ and } \sum_{i \in [n]} x_i = \omega([n])\}. \tag{3}$$

A wall in the S_n fan corresponding to $i \perp\!\!\!\perp j \mid K$ is missing in the normal fan of P defined by ω as above if and only if $\omega(Ki) + \omega(Kj) = \omega(Kij) + \omega(K)$. In particular, a coarsened S_n fan is polytopal if and only if the corresponding semigraphoid is submodular.

Proof of Lemma 2.3. Let $P = \{x \in \mathbb{R}^n : Ax + b \geq 0\}$, where A is a $k \times n$ matrix whose rows positively span $\mathbb{R}^n$ and $b \in \mathbb{R}^k$ is a column vector. Suppose the polytope P is non-empty and that all inequalities are tight but possibly redundant. Let C be the cone in $\mathbb{R}^n \times \mathbb{R}$ generated by the rows of the concatenated matrix $[A|b]$. The row (a_i, b_i) of $[A|b]$ is called a *lift* of the vector a_i. The dual cone $C^*:=\{u : u^T v \geq 0 \ \forall v \in C\}$ is generated by $\{(x, 1) : x \in P\}$. The projection of proper faces of C onto $\mathbb{R}^n$ forms the inner normal fan of P. Since all inequalities are assumed to be tight, all lifted vectors (a_i, b_i) lie on the boundary of C. All vectors in the dual cone C^* have positive last coordinates, so all proper faces of C are on the lower hull of C. In particular, if a vector (a, b) lies on the boundary of C, then $(a, b + \varepsilon)$ does not lie on the boundary of C for any $\varepsilon > 0$. See [DLRS10, §2.5] for more details on this construction.

Let F be a polytopal fan which coarsens the S_n fan. Every cone in F contains a line in direction $(1, 1, \ldots, 1)$ and is generated by this line together with some $0/1$ vectors. From the discussion above, there exists a lift (real valued function) ω on the set of rays $\{e_I \mid \varnothing \neq I \subset [n]\} \cup \{-e_{[n]}\}$ such that the faces of the cone C spanned by the lifted rays (vectors $(r, \omega(r))$ where r is a ray of F) project precisely onto the cones of F and every lifted ray is on the boundary of C. Since all cones in F contain the line $(1, 1, \ldots, 1)$, it suffices to consider lifts ω such that $\omega(e_{[n]}) = -\omega(-e_{[n]})$. Such lifts can be identified with functions on $2^{[n]}$ with value 0 on $\varnothing$. We will show that ω is submodular.

For any $I, J \subset [n]$, the vectors $e_I, e_{I\cap J}, e_{I\cup J}$ lie in a common cone in the S_n fan. Since F coarsens the S_n fan, they also lie in a common cone in F. Similarly $e_J, e_{I\cap J}, e_{I\cup J}$ lie in a common cone of F. First, consider the case when e_I and e_J are lifted to the same proper face of C. Then this cone also contains $e_{I\cap J}$ and $e_{I\cup J}$. Since we assumed that all lifted vectors lie on the boundary, hence a proper face, of C, and ω is linear on this face, we must have that $\omega(e_I) + \omega(e_J) = \omega(e_{I\cap J}) + \omega(e_{I\cup J})$.

Now suppose that e_I and e_J are not lifted to the same proper face of C. Then ω is not linear on the vectors $e_I, e_J, e_{I\cap J}$, and $e_{I\cup J}$. We must then have that $\omega(e_I) + \omega(e_J) > \omega(e_{I\cap J}) + \omega(e_{I\cup J})$, because $\omega(e_I) + \omega(e_J) < \omega(e_{I\cap J}) + \omega(e_{I\cup J})$ would imply that

$$(e_{I\cap J} + e_{I\cup J}, \omega(e_{I\cap J}) + \omega(e_{I\cup J})) > (e_I + e_J, \omega(e_I) + \omega(e_J)),$$

contradicting the fact that $e_{I\cap J}$ and $e_{I\cup J}$ are lifted to the same cone in the lower hull of C.

For the converse, suppose ω is a submodular function on $2^{[n]}$ with $\omega(\varnothing) = 0$ and consider the lift of e_I to $\omega(I)$ for each $I \subseteq [n]$ and $-e_{[n]}$ to $-\omega([n])$. Let F be the projection of the lower hull of the lifted cone C. The submodularity inequality $\omega(e_I) + \omega(e_J) \geq \omega(e_{I\cap J}) + \omega(e_{I\cup J})$ ensures that whenever I and J are lifted to the same cone in the lower hull of C, then so are $I \cap J$ and $I \cup J$. In other words, whenever a cone of F contains both, e_I and e_J, then it must also contain both, $e_{I\cap J}$ and $e_{I\cup J}$, showing that F is a coarsening of the S_n fan.

Now suppose that the coarsened S_n fan F is polytopal defined by a submodular function ω as above. The wall corresponding to the pair of adjacent permutations in (1) is not a cone in the fan F if and only if the two adjacent maximal cones are contained in the same cone of F. In particular, this happens if and only if e_{Ki} and e_{Kj} are in the same cone where $K = \{a_1, \ldots, a_k\}$. This is equivalent to the condition that (2) is attained at equality.

Let P be the polytope defined by (3). Its inner normal fan is obtained by lifting the rays $-e_I$ to height $\omega(I)$ for non-empty $I \subset [n]$ and $e_{[n]}$ to height $-\omega([n])$. This is the negation of the fan F, which is obtained by lifting e_I to $\omega(I)$ and $-e_{[n]}$ to $-\omega([n])$. This shows that F is the outer normal fan of P. □

Appendix C. Dictionary

The statements or data in each row are equivalent.

CI relations	**fans**	**polytopes**
CI relation $i \perp\!\!\!\perp j \mid K$ where $i, j \in [n]$, $K \subset [n] \setminus \{i, j\}$	the set of walls in the S_n fan of the form $\sigma \vert i\, j \vert \tau$ where σ and τ are permutations of K and $[n] \backslash Kij$ respectively	the set of edges of a permutohedron connecting two permutations of the form $\sigma \vert i \vert j \vert \tau$ and $\sigma \vert j \vert i \vert \tau$ where σ and τ are permutations of K and $[n] \backslash Kij$, respectively
a collection of CI relations that satisfy the semigraphoid axioms	removing the walls in the S_n fan corresponding to the independence relations gives a fan	the set of edges of the permutohedron corresponding to the independence relations satisfies the square and hexagon axioms [MPS+09]
a semigraphoid that arises from a submodular function	a coarsening of S_n fan that is *polytopal* or *regular*	there is a generalized permutohedron that realizes contraction of edges in the permutohedron corresponding to the CI relations
a union of dependence relations of a semigraphoid	a common refinement of fans	a Minkowski sum of polytopes (if the semigraphoid is submodular)

Acknowledgement

CU was partially supported by DARPA (DARPA-SN-16-37), ONR (N00014-16-S-BA10) and the Austrian Science Fund (FWF) Y 903-N35. JY was partially supported by the US NSF grant DMS #1600569. We are grateful to the anonymous referees for very helpful comments on earlier versions of this paper.

References

[AMP97] S. A. Andersson, D. Madigan, and M. D. Perlman, *A characterization of Markov equivalence classes for acyclic digraphs*, The Annals of Statistics **25** (1997), no. 2, 505–541.

[CD06] M. P. Carr and S. L. Devadoss, *Coxeter complexes and graph-associahedra*, Topology and its Applications **153** (2006), no. 12, 2155–2168.

[Chi02a] D. M. Chickering, *Learning equivalence classes of Bayesian-network structures*, Journal of Machine Learning Research **2** (2002), no. 3, 445–498.

[Chi02b] D. M. Chickering, *Optimal structure identification with greedy search*, Journal of Machine Learning Research **3** (2002), 507–554.

[Chi95] D. M. Chickering, *A transformational characterization of equivalent Bayesian network structures*, Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, 1995, pp. 87–98.

[CHS16] J. Cussens, D. Haws, and M. Studený, *Polyhedral aspects of score equivalence in Bayesian network structure learning*, 2016. to appear in *Mathematical Programming, Series A*.

[Dev09] S. L. Devadoss, *A realization of graph associahedra*, Discrete Mathematics **309** (2009), no. 1, 271–276.

[DLRS10] J. A. De Loera, J. Rambau, and F. Santos, *Triangulations*, Algorithms and Computation in Mathematics, vol. 25, Springer-Verlag, Berlin, 2010.

[DSS09] M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on Algebraic Statistics*, Oberwolfach Seminars, vol. 39, Birkhäuser Verlag, Basel, 2009.

[DX10] M. Drton and H. Xiao, *Smoothness of Gaussian conditional independence models*, Algebraic methods in statistics and probability II, 2010, pp. 155–177.

[Fuj05] S. Fujishige, *Submodular Functions and Optimization*, Second, Annals of Discrete Mathematics, vol. 58, Elsevier, Amsterdam, 2005.

[HLS12] Raymond Hemmecke, Silvia Lindner, and Milan Studenỳ, *Characteristic imsets for learning bayesian network structure*, International Journal of Approximate Reasoning **53** (2012), no. 9, 1336–1349.

[HMS+08] R. Hemmecke, J. Morton, A. Shiu, B. Sturmfels, and O. Wienand, *Three counter-examples on semi-graphoids*, Combinatorics, Probability and Computing **17** (2008), no. 2, 239–257.

[JSGM10] T. Jaakkola, D. Sontag, A. Globerson, and M. Meila, *Learning Bayesian network structure using LP relaxations*, Proceedings of the 13th international conference on artificial intelligence and statistics, 2010, pp. 358–365.

[Lau96] S. L. Lauritzen, *Graphical Models*, Oxford Statistical Science Series, vol. 17, The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.

[LM07] R. Lněnička and F. Matúš, *On Gaussian conditional independent structures*, Kybernetika (Prague) **43** (2007), no. 3, 327–342.

[Mee95] C. Meek, *Causal inference and causal explanation with background knowledge*, Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, 1995, pp. 403–410.

[Mee97] C. Meek, *Graphical models: Selecting causal and statistical models*, 1997. PhD thesis, Carnegie Mellon University.

[MPS+09] J. Morton, L. Pachter, A. Shiu, B. Sturmfels, and O. Wienand, *Convex rank tests and semigraphoids*, SIAM Journal on Discrete Mathematics **23** (2009), no. 3, 1117–1134.

[MSUW17] L. Matejovicova, L. Solus, C. Uhler, and Y. Wang, *Restricted permutation search for causal inference with background knowledge*, 2017. arXiv:1702.03530.

[Mur03] K. Murota, *Discrete Convex Analysis*, SIAM Monographs on Discrete Mathematics and Applications, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003.

[Oxl11] James Oxley, *Matroid theory*, Second Edition, Oxford Graduate Texts in Mathematics, vol. 21, Oxford University Press, Oxford, 2011. MR2849819

[Pea88] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, The Morgan Kaufmann Series in Representation and Reasoning, Morgan Kaufmann, San Mateo, CA, 1988.

[PRW08] A. Postnikov, V. Reiner, and L. Williams, *Faces of generalized permutohedra*, Documenta Mathematica **13** (2008), 207–273.

[RU14] G. Raskutti and C. Uhler, *Learning directed acyclic graphs based on sparsest permutations*, 2014. arXiv:1307.0366.

[Sha98] R. D. Shachter, *Bayes-ball: Rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams)*, Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1998, pp. 480–487.

[SHHL12] M. Studený, D. Haws, R. Hemmecke, and S. Lindner, *Polyhedral approach to statistical learning graphical models*, Harmony of gröbner bases and the modern industrial society: the 2nd crest-sbm international conference (T. Hibi ed.), 2012, pp. 346–372.

[SL14] Kayvan Sadeghi and Steffen Lauritzen, *Markov properties for mixed graphs*, Bernoulli **20** (2014), no. 2, 676–696. MR3178514

[Stu05] Milan Studený, *Probabilistic conditional independence structures*, Information Science and Statistics, Springer, London, 2005. MR3183760

[Stu92] M. Studenỳ, *Conditional independence relations have no finite complete characterization*, Information Theory, Statistical Decision Functions, Random Processes **B** (1992), 377–396.

[Sul09] S. Sullivant, *Gaussian conditional independence relations have no finite complete characterization*, Journal of Pure and Applied Algebra **213** (2009), no. 8, 1502–1506.

[SVH10] M. Studený, J. Vomlel, and R. Hemmecke, *A geometric view on learning Bayesian network structures*, International Journal of Approximate Reasoning **51** (2010), 578–586.

[TK05] M. Teyssier and D. Koller, *Ordering-based search: A simple and effective algorithm for learning Bayesian networks*, Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, 2005, pp. 584–590.

[URBY13] C. Uhler, G. Raskutti, P. Bühlmann, and B. Yu, *Geometry of the faithfulness assumption in causal inference*, The Annals of Statistics **41** (2013), no. 2, 436–463.

[VP90] T. Verma and J. Pearl, *Equivalence and synthesis of causal models*, Proceedings of the 6th Annual Conference on Uncertainty in Artificial Intelligence, 1990, pp. 255–270.

[Zie95] G. M. Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, vol. 152, Springer-Verlag, New York, 1995.

School of Mathematics, University of Bristol, Bristol, BS8 1TW, UK

E-mail address: `fatemeh.mohammadi@bristol.ac.uk`

Department of Electrical Engineering & Computer Science, and Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge MA, USA

E-mail address: `cuhler@mit.edu`

School of Mathematics, Georgia Institute of Technology, Atlanta GA, USA

E-mail address: `charles.wang@gatech.edu`

School of Mathematics, Georgia Institute of Technology, Atlanta GA, USA

E-mail address: `jyu@math.gatech.edu`